



```

163 CCGAAG.....GATATTATAGACTCGTTTAAAGGGGGCATCGCGAG 203
   : : : : : : : : : : : : : : : : : : : : : : : : : :
59 GluArgAsnArgProGluTyrIleAspLeuLeuLysSerLeuIleLysIy 75
   : : : : : : : : : : : : : : : : : : : : : : : : : :
204 TGACCTGATAGATATCGAAGACGAGCTACAGCAGCAATACTCCGCC 253
   : : : : : : : : : : : : : : : : : : : : : : : : : :
75 sGlyGlnValGluLeuValAlaGlyPheTyrGluProIleLeuValA 92
   : : : : : : : : : : : : : : : : : : : : : : : : : :
254 TCCTCCGCTTAGCAGAGTAGAAGACAAAGTTCAGAGATAGGGAAGTT 303
   : : : : : : : : : : : : : : : : : : : : : : : : : :
92 lailePro.....GluGluAsp 97
   : : : : : : : : : : : : : : : : : : : : : : : : : :
304 AAGGAAGAGCTCTCGAGCTTCT..... 327
   : : : : : : : : : : : : : : : : : : : : : : : : : :
98 ArgValGluGlnIleLysLeuSerLysGlyTyrPalaArgLysMetGlyTy 114
   : : : : : : : : : : : : : : : : : : : : : : : : : :
328 ....CCAAGGATTCGTGCGCAGAGCTCGCTATGACCCGATATACC 373
   : : : : : : : : : : : : : : : : : : : : : : : : : :
114 rGluAlaArgGlyLeuTrpLeuThrGluArgValTrpGluProGluLeuV 131
   : : : : : : : : : : : : : : : : : : : : : : : : : :
374 CTGCCTACTGAAGCAACGGTTATGATCTATTCGCCGACGGGAG 423
   : : : : : : : : : : : : : : : : : : : : : : : : : :
131 allYsThrLeuArgGluAlaGlyIleGluTyrValIleLeuAsp..... 145
   : : : : : : : : : : : : : : : : : : : : : : : : : :
424 CGGATGCTTTCTCAGCTCATCTCACTCGCG.....ATAAGCCAATTAA 470
   : : : : : : : : : : : : : : : : : : : : : : : : : :
146 .....AspTyrHisPheMetSerAlaGlyLeuSerLysGluI 158
   : : : : : : : : : : : : : : : : : : : : : : : : : :
471 ACCGCTCTATCCACCTTATTAAGGCCCAAGGAAAGCGCTTTAGCT 520
   : : : : : : : : : : : : : : : : : : : : : : : : : :
158 uLeuPheTrpProTyrThrGluAsnGlyGlyGluAlaIleValValP 175
   : : : : : : : : : : : : : : : : : : : : : : : : : :
521 AC.....ATCAGTATCTCTCTGCTCTCAGGGAGCTTAGG 555
   : : : : : : : : : : : : : : : : : : : : : : : : : :
175 heProIleAspGluLysLeuArgTyrLeuIleProPheArgProValasn 191
   : : : : : : : : : : : : : : : : : : : : : : : : : :
556 AAGCGGATAAGCTCGCTTT.....GAAGTAAGT 587
   : : : : : : : : : : : : : : : : : : : : : : : : : :
192 GluThrLeuGluTyrLeuHisSerLeuAlaAspGluAspGluSerLysVa 208
   : : : : : : : : : : : : : : : : : : : : : : : : : :
588 AAGCGTAAGGCAGCTCAAGACATCGAAGCGTACCGTTTGGGTGGCCG 637
   : : : : : : : : : : : : : : : : : : : : : : : : : :
208 lAlaVal...PheHisAspAspGlyGlyLysPheGlyAlaTrpProGlyT 224
   : : : : : : : : : : : : : : : : : : : : : : : : : :
638 TGAACACGCTGTATGCTCGGCATCGGAAGGCTTCTCTTATGATCTCT 687
   : : : : : : : : : : : : : : : : : : : : : : : : : :
224 hrHisGluLeuValTyr..... 229
   : : : : : : : : : : : : : : : : : : : : : : : : : :
688 AAGAAGTGGCGAGCTGATA.....GA 710
   : : : : : : : : : : : : : : : : : : : : : : : : : :
230 .....GluArgGlyTyrLeuLysGluPhePheAspArgIleSerAs 244
   : : : : : : : : : : : : : : : : : : : : : : : : : :
711 GGAAGGACACATCTTCTATACGACCGGATATAGATTATGCG. 759
   : : : : : : : : : : : : : : : : : : : : : : : : : :
244 pasPlysIleAsnLeuMetLeuTyrSer.....GluTyrLeuSerL 258
   : : : : : : : : : : : : : : : : : : : : : : : : : :
760 .....TATAGGACATTGACGGCTAC...AGA 783
   : : : : : : : : : : : : : : : : : : : : : : : : : :
258 ysPheArgProLysLeuValTyrLeuProIleAlaSerTyrPheGlu 274
   : : : : : : : : : : : : : : : : : : : : : : : : : :
784 ATGAGTGTGAGGGATA.....TTAGAGGT 809
   : : : : : : : : : : : : : : : : : : : : : : : : : :
275 MetSerGluTrpSerLeuProAlaArgGlnAlaLysLeuPhePheGluPh 291
   : : : : : : : : : : : : : : : : : : : : : : : : : :
810 TATAGACGCTCACTCGAGCTGTGCTTCCCTCAGAGCTGAGGCACA 859
   : : : : : : : : : : : : : : : : : : : : : : : : : :
291 elleLysLysLeu...LysGluLeuAsnLeuPheGluLysTyrArg... 305
   : : : : : : : : : : : : : : : : : : : : : : : : : :
860 GTGGAAGGAGCTCTACTAGGACTTCGAGTTGGCCACCATAGAGGC 909
   : : : : : : : : : : : : : : : : : : : : : : : : : :
306 .....IlePheValArgGlyIleIleTrp.....LysAsn 315
   : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

910 TTGAGGATATGGAGAGACGAAGGGAAC.....GCAAGACTTAA 950
   : : : : : : : : : : : : : : : : : : : : : : : : : :
316 Phe...LeuTyrLysTyrProGluGlyAsnTyrMetHisLysArgMetLe 331
   : : : : : : : : : : : : : : : : : : : : : : : : : :
951 TATGCTGTCTCAATATAGGGGGGAACTCGCCCTTTAGCCGAGAACA 1000
   : : : : : : : : : : : : : : : : : : : : : : : : : :
331 uMetLeuSer.....LysLeuLeuArgAsnAsn. 340
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1001 GCATGCAAGGGGATGGAGCCCCCTCCCTGAGAGGAGGCTGATGCCTTC 1050
   : : : : : : : : : : : : : : : : : : : : : : : : : :
341 .....ProThrAlaArgIlePheValLeu 348
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 CGGCGCATATATAACGAT.....TGGAGGGGTGAAATGGG 1086
   : : : : : : : : : : : : : : : : : : : : : : : : : :
349 ArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPheGly 362
   : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_40:AMY1_DICTH

seq_documentation_block:
ID AMY1_DICTH STANDARD; PRT; 685 AA.
AC P09961;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).
GN AMYA.
OS Dictyoglomus thermophilum.
OC Bacteria; Dictyoglomus group; Dictyoglomus.
OX NCBI_TaxID=14;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H-6-12.
RX MEDLINE=8822597; PubMed=2453362;
RA Fukusumi S., Kamazono A., Horinouchi S., Beppu T.:
RT "Cloning and nucleotide sequence of a heat-stable amylase gene from
RT an anaerobic thermophile, Dictyoglomus thermophilum."
RL Eur. J. Biochem. 174:15-21(1988).
CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OLIGOMERS
CC APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED
CC DECREASE IN THE AMOUNTS OF MALTOPIROSE, MALTOSE AND GLUCOSE IN
CC PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE: THE OPTIMUM
CC TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES
CC CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO
CC DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80
CC DEGREES CELSIUS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X07896; CAA30735.1; -.
CC DR PIR: S00628; ALDYAT.
CC DR InterPro: IPR004300; Glyco_hydro_57.
CC DR Pfam: PF03065; Glyco_hydro_57; 1.
CC KW Hydroxylase; Glycosidase; Carbohydrate metabolism; Multigene family.
CC FT INIT_MET 0
CC SQ SEQUENCE 685 AA; 81060 MW; C10941C8A508C404 CRC64;

```

```

alignment_scores:
  Quality: 157.50      Length: 315
  Ratio: 0.960        Gaps: 15
  Percent Similarity: 52.063      Percent Identity: 22.540

```

```
alignment_block:
US-09-886-400-3 x AMYL_DICTH ..
Align seg 1/1 to: AMYL_DICTH from: 1 to: 685

22 GCAACCTCCAGTATCGCGAAATCCCAAGAGCGAAATCCCAAGAGTCAT 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 GlyAsnPheAspPheVal.....IleGluArgAlaIty 27

72 ACAGAGCATACATCCCGAGTCATCGACACACTGATTAAAGAA...GAAA 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
27 rGluMetSerTyrLysProLeuIleAsnPhePheLeuLysProAsp 44

119 TTCTTTGGGCTCAACATAACGGCTATACCTTAAAGTTCTCTCCGGAAG 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 heProIleAsnValHisPheSerGlyPheLeuLeuLeuTrpLeuGluLys 60

169 .....GATATTATACCTCTTAAGGGCGCATCGCGAGTGACCT 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 AsnHisProGluTyrPheGluLysLeuLysIleMetAlaGluArgGly 77

210 GATAGAGATAATCGAAGAGCTACACGACGCAATACCTCCCTCTCTCC 259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
77 nleGluPheValSerGlyGlyPheTyrGluProIleLeuLeuIleP 94

260 CGCTTAGCAGATAGACACACAGCTTCAGACAGATAGGGAAGTTAAGAA 309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
94 roAspLysAspLysValGlnGlnIleLysLysLeuAsnLysTyrIleTyr 110

310 GAGCTCTCGAGCTTCTCCAAAGGATTCGTGGCTCGCAGAGCTCGCCTA 359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 AspLysPheGlyGlnThrProLysGlyMetTrpLeuAlaGluArgValTr 127

360 TGACCCGATAATCCCTGTCATCTACTGAAGACACACGTTATGATGATCTAT 409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 pGluProHisLeuValLysTyrIleAlaGluAlaGlyIleGluTyrVal 144

410 TCGCCGACGGGAGGAGTGCTTTCTCAGCTCATCTCAACTCGCGGATA 459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
144 alValAsp...AspAlaHisPheSerValGlyLeu..... 155

460 AAGCCCAATTAACCGCTCTATCCACACCTTATAAAGGCCCAAGGAAAA 509
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 .....LysGluGluAs 159

510 GCCTTTAGTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAGG 559
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 pLeuPheGlyTyr.....TyrLeuMetGluGluGlnGlyTyrLysLeuA 174

560 CGATAAAGCTCGTTTGAAGTAAAGTAAAGCTAAAGCGAGTCAAGAC 609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 la.....ValPheProIleSerMetLysLeuArgTyrLeu..... 185

610 ATCGAAGCCCTACCCCTTTGGTGGCGGTGAACACGGCTGTAATGCTCG 659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ..... 185

660 CATCGGAAGCTTCTCTTATGATCTTAAAGTGGCGAGCTGGATA. 708
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 .....IleProPheAlaAspProGluGlnThrIleThrTyrLeuA 199

709 .....GAGGACAGGACAACTTCTTATACGCGACCGAT 744
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 spLysPheAlaSerGluAspLysSerLysIleAlaLeuLeuPheAsp 215

745 ATAGAGTTCATTGGC.....TATAGGCACATTGCAGGCTA 779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 GlyGluLysPheGlyLeuTrpProAspThrTyrArgThrVal.....Ty 230

780 CAGAGTGAAGTCTGAGGATATTATAGGTT...ATAGACGAGCTCAACT 826
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
230 rGlu.....GluGlyTrpLeuGluThrPheValSerLysIleLysG 244
```

```
827 CGGAAGTGTGCTTCCCTCA.....GAG 849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 LuasPheLeuLeuValThrProValAsnLeuTyrThrMetGlnArg 260

850 CTGAAGCAGTGGAGGAGCTCTACTTACGAGCTTCGAGTTGG 894
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 ValLysProLysGlyArg...IleTyrLeuProThrAlaSerTyr 274

seq_name: SwissProt_40: AMYA_PYRAB

seq_documentation_block:
ID AMYA_PYRAB STANDARD; PRT; 655 AA.
AC Q9V298; 2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-amylase (EC 3.2.1.1).
GN AMYA OR PAB0118.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID:29292;
RN [1]
RC STRAIN=ORSAY;
RP SEQUENCE FROM N.A.
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
linkages in oligosaccharides and polysaccharides.
CC -|- PATHWAY: POLYSACCHARIDE DEGRADATION.
CC -|- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ248283; CAB49100.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
DR Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
SQ SEQUENCE 655 AA; 77296 MW; 7F6F920B1A00EECE CRC64;

alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent Identity: 22.444

alignment_block:
US-09-886-400-3 x AMYA_PYRAB ..
Align seg 1/1 to: AMYA_PYRAB from: 1 to: 655

73 GAGAAGGCATACATCCAGTCATCGACACACTTAAAGAGAGAATTC 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 GluLysAlaTyrArgProPheLeuGluIleLeu.....GluGluTyrPr 42

123 T.....TTTGGCTCAACATAACGGCTATACCTTAAAGTTCTCTCC 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 oAsnMetLysValAlaIleHisSerGlyIleLeuValGluTrpLeuG 59

164 CGAAG.....GATATTATAGACCTCTTAAAGGGCGCATCGCGAGT 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 LuGluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75

205 GACCTGATAGAGATAATCGAAGCAGCTACACGACGCAATCTCCCTCC 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 GlycInValGluIleValValAlaGlyPheTyrGluProValLeuAla 92
```

```
255 CCTCCCTTAGCAGATAGAACCAACCAAGTTTCAGAGATAGAGGAGTTA 304
    :||| :| :||| :||| :||| :||| :||| :||| :||| :|||
92 aileProLysGluAspArgLeuGluGlnIleTyrLeuLeuLysGluTrpA 109
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
305 AGGAAGAGCTTCGAGCTTCCTCCAAAGGATTCGTGGCTCCAGAGCTC 354
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
109 laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
355 GCCTATGACCCGATATCCCTGCCATCTAGTACGAGGACGAGGTTATGAT 404
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
125 ValTrpGlnProGluLeuValThrLeuArgGluAlaGlyIleGly 141
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
405 TCTATTCGCGAGCGGGAGCGGATGCTTTCTCAGCTCATCTCACTCGG 454
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
141 rValValValAsp...AspTyrHisPheMetSerA 152
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 CG...ATAAGCCCAATTAACCGCTCTATCCACACCTTTATAAGGCCAA 501
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
152 laGlyLeuSerLysAspGlnLeuPheTrpProTyrTyrThrGluAspGly 168
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 AGGGA...AGCGCTTAGGTAC... 522
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
159 GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeu 185
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
523 .....ATCAGCTATCTCTGTGCTCAGGG 547
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
185 eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAla 202
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
548 AGCTTAGGAGCGCATAAAGCTGTTTTCAGGTAAGTAAGCTAAAG 597
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
202 erGluAspGluSerLysValAlaValPheHis..... 212
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
598 GCAGTCAAGACATCGAAGCGCTACCGCTTGGTGGCGCGTGAACAGGC 647
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 .....AspAspGlyGluLysPheGlyIleTrp..... 221
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
648 TGTATGCTCGCATCGAAGGCTTCTCTTATGAATCTTAAGAAAGTGG 697
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
222 .....ProMet.....Thr 225
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
698 CGAGCTGATAGAGGACAAAGGAC..... 720
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 yrGluTrpValTyrGluLysGlyTrpLeuArgGluPheAspArgVal 241
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
721 .....ACATCTCTATAGCGCACCATATAG... 750
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
242 SerSerAspGluAlaIleAsnIleMetLeuTyrSerGluTyrLeuGln 258
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
751 .....TTCATTGGCTATAGGACATTGCAGGCTAC..... 780
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
258 sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGlu 275
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
781 .....AGATGAGTGTGAGGA 798
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
275 etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
799 TTATTAGAGGTATAGACGAGCTCAACTCGGAAGTGTGCTTCCCTCAGA 848
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
291 PheValGluLysLeuLysGluLeuAsn..... 299
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
849 GCTGAAGCACAGTGAAGGAGCTCTACTTACGACCTCGAGTTGGGCAC 898
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
300 .....MetPheGluArgTyrArgValPheValArgGlyGlyIleTrp... 313
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
899 CAGATAAGAGCTTGAGATATGGAGAGGAGGACGAAGGGRAC..... 939
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
314 .....LysAsnPhe...PheTyrLysTyrProGluAlaAsnTyrMeth 327
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
940 GCAGACTTAATATGCTCTCTACAAATATGAGGGCGAAGTCTGCCCTTT 989
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
328 LysArgMetLeuMetLeuSer.....ArgLeuLe 337
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
990 AGCGGAGAACAGCGATGCAAGGGATGGAGCCCTCCCTGAGAGGAGGC 1039
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
337 urAspAsnProSerAlaArgPhe..... 346
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1040 TGGATGCTTCCCGGCGATATATACGAT.....TGGAGGGGTGAAAT 1083
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
347 .....ValLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPhe 361
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1084 GGG 1086
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
362 Gly 362
```

seq\_name: SwissProt\_40:AMYA\_METJA

```
seq_documentation_block:
ID AMYA_METJA STANDARD; PRT; 467 AA.
AC Q59006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative alpha-amylase (EC 3.2.1.1).
GN MJ1611.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Adams M.D., Reich C.I.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Merrick J.M., Glodek A.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Fuhmann J.L., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Babor P.W., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii." 1996.
RL Science 273:1058-1073(1996).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC linkages in oligosaccharides and polysaccharides.
CC -I- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U67601; AAB99631.1;
CC TIGR; M01611;
CC InterPro; IPR004300; Glyco_hydro_57.
CC Pfam; PF03065; Glyco_hydro_57; 1.
CC KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;
CC Complete proteome.
CC SEQUENCE 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;
```

alignment\_scores:  
Quality: 141.50 Length: 377  
Ratio: 0.741 Gaps: 24  
Percent Similarity: 50.663 Percent Identity: 22.812

alignment\_block:  
US-09-886-400-3 x AMYA\_METJA ..  
Align seg 1/1 to: AMYA\_METJA from: 1 to: 467

40 GAAATCCCAAAGAGCGAAATCCCAAGGTCATAGAGAAGGCATACATCC 89





```

FT   INIT MET      0
SQ   SEQUENCE 648 AA; 76178 MW; BF7A495F084E0FB1 CRC64;

alignment_scores:
  Quality: 141.50      Length: 391
  Ratio: 0.773         Gaps: 23
  Percent Similarity: 46.803 Percent Identity: 22.762

alignment_block:
US-09-886-400-3 x AMYA_PYRFU
Align seg 1/1 to: AMYA_PYRFU from: 1 to: 648

73  GAGAGGATACATCCAGTCTCGAGACATGATTAAAGAGAATTC 122
|||||:||||| ||| |||||:||||| ||| |||||:|||||
28  GlulysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 42
123 T.....TTTGGGCTCAACATACAGGGCTATACCTTAAGTTCTCC 163
|||||:||||| ||| |||||:||||| ||| |||||:|||||
42  oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 59
164 CG.....AAGGATATTATAGACTCGTTAAAGGGGCGATCGGAGT 204
|||||:||||| ||| |||||:||||| ||| |||||:|||||
59  lAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 75
205  GACCTGATAGATAATCGGAACAGAGCTACACGACGCAATCTCCCTC 254
|||||:||||| ||| |||||:||||| ||| |||||:|||||
76  GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAlaLase 92
255  CTTCCCGCTTAGCAGATAGACCAAGTTCAGAGAGATAGGAAGTTA 304
|||||:||||| ||| |||||:||||| ||| |||||:|||||
92  rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 106
305  AGGAA.....GAGCTCTTCGAGCTTCTCCAAAGGATTCGGTGCCA 348
|||||:||||| ||| |||||:||||| ||| |||||:|||||
106  ySglutPAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 122
349  GAGCTCGCTATACCGGATATCCCTGCCATCTACGAGCAACGGTTA 398
|||||:||||| ||| |||||:||||| ||| |||||:|||||
123  GluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGlyTl 139
399  TGAGTATCTATTGCGCGAGCGGGGAGCGATGCTTTTCTCAGCTCATCTCA 448
|||||:||||| ||| |||||:||||| ||| |||||:|||||
139  eAspTyrValIleValAsp.....AspTyrHisPheM 150
449  ACTCGCGGATAAAGCCCAATTAACCGCTCTAT...CCACACCTTATAAG 495
|||||:||||| ||| |||||:||||| ||| |||||:|||||
150  etSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrTrpThrGlu 166
496  GCCCAAAGGAA.....AAGCGCTTAGGTA 521
|||||:||||| ||| |||||:||||| ||| |||||:|||||
167  AspGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 182
522  CACAGATATCTCTTGCTCGAGGAGCTTAGAAGGCGATAAAGCTC. 570
|||||:||||| ||| |||||:||||| ||| |||||:|||||
183  .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 197
571  .....GTTTGTAGGTAAGCTAACGCTAAAGCGAGTC.....AAA 606
|||||:||||| ||| |||||:||||| ||| |||||:|||||
197  euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 213
607  CACATCAAGCCGTACCGCTTTGGGTGGCGGTGAACACGCGCTGCTATGCT 656
|||||:||||| ||| |||||:||||| ||| |||||:|||||
214  AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 229
657  CGGATCGGAGGCTTCTCTTATGAATCTTAAGAAAGTGGCGAGTGA 706
|||||:||||| ||| |||||:||||| ||| |||||:|||||
230  .....GluLysGlyTrpL 234
707  TA.....GAGCAACAGGACACATCTTC 729
|||||:||||| ||| |||||:||||| ||| |||||:|||||
234  euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 250

```

```

730  CTATACGGCACCGCATATAGAG.....TTCATTGGCTATAG 764
|||||:||||| ||| |||||:||||| ||| |||||:|||||
251  LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 267
765  GGACATTGCGGCTACAGAATGAGTGTTCGAGGATATTAGAGGTTATAG 814
|||||:||||| ||| |||||:||||| ||| |||||:|||||
267  uProIleAlaSerTyr.....PheGluMet.... 275
815  ACGAGCTCAACTCGGAACGTGCTCCCTCC..... 843
|||||:||||| ||| |||||:||||| ||| |||||:|||||
276  .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 288
844  .....TCAGAGCTGAAGCACAGTGA.....AGGA 869
|||||:||||| ||| |||||:||||| ||| |||||:|||||
289  ValGluPheValAsnGluLysValLysGlyIlePheGluLysTyrAr 305
870  GCTCTACTTACGAGCTTCGAGTTGGCACCAGATAAGAGCTTGAGGATAT 919
|||||:||||| ||| |||||:||||| ||| |||||:|||||
305  gValPheValArgGlyGlyLeuTrp.....LysAsnPhe...PheT 318
920  GGAGAGAGGACGAAGGAAC.....GCAAGACTTAATATGCTGTCC 960
|||||:||||| ||| |||||:||||| ||| |||||:|||||
318  yrlLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 334
961  TACAATATGAGGGCGGAACCTCCCTTTTAGCCGAGAACAGCGATGCAAG 1010
|||||:||||| ||| |||||:||||| ||| |||||:|||||
335  LysLeuValArgAsnAsn..... 340
1011  GGGATGGAGCCCTCCCTGAGAGGAGCGTGGATGCTTCGGGCGGATAT 1060
|||||:||||| ||| |||||:||||| ||| |||||:|||||
341  .....ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 352
1061  ATAACGAT.....TCGAGGGGT 1077
|||||:||||| ||| |||||:||||| ||| |||||:|||||
352  ysAsnAspAlaTyrTrpHisGly 359
seq_name: SwissProt_40:MALQ_PYRKO
seq_documentation_block:
ID  MALQ_PYRKO STANDARD; PRT: 653 AA.
AC  O32450;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE  (Disproportionating enzyme) (D-enzyme).
OS  Pyrococcus kodakaraensis.
OC  Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX  NCBI_TaxID=69014;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=KOD1;
RA  Tachibana Y., Fujiwara S., Takagi M., Imanaka T.;
RT  "Cloning and expression of the 4-alpha-glucanotransferase gene from
RT  the hyperthermophilic archaeon Pyrococcus sp. KOD1, and
RT  characterization of the enzyme.";
RL  J. Ferment. Bioeng. 83:540-548(1997).
CC  -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
CC  to a new 4-position in an acceptor, which may be glucose or (1,4)-
CC  alpha-D-glucan.
CC  -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  EMBL; D87907; BAA22062.1;
CC  InterPro; IPR004300; Glyco_hydro_57.
DR

```

DR Pfam; PF03065; Glyco\_hydro\_57; 1.  
 KW Transferase; Glycosyltransferase; Carbohydrate metabolism.  
 SQ SEQUENCE 653 AA; 76664 MW; B0C4695613F29219 CRC64;

alignment\_scores:  
 Quality: 121.50 Length: 386  
 Ratio: 0.653 Gaps: 21  
 Percent Similarity: 48.187 Percent Identity: 21.244

alignment\_block:

US-09-886-400-3 x MALQ\_PYRKO ..

Align seg 1/1 to: MALQ\_PYRKO from: 1 to: 653

```

73 GAGAGGATACATCCAGTCATCGAGACATGATTAAAGAGAAATCC 122
|||||:||||| ||| :|||:||||| |||:||||| |||
28 GluArgSerTyrArgProPheMetGluThrLeu.....GluGluTyrPr 42
123 T.....TTGGGCTCAACATAACGGGTATACCTTAAAGTTCCTC. 162
|||||:||||| :||| |||:||||| |||:|||||
42 OsnMetLysValAlaValHisTyrSer-GlyProLeuLeuGluTrpIleA 59
163 .....CCGAGGATATTATAGACCTCGTTAAAGGGGCATCGCAGT 204
|||:||||| ||| :|||:||||| |||:||||| |||
59 rGAspAsnLysProGluHisLeuAspLeuLeuArgSerLeuValLysArg 75
205 GACCTGATAGAGATAATCGGACAGAGCTACACGACGCAATACTCCCCCT 254
|||||:||||| :||| |||:||||| |||:||||| |||
76 GlyGlnLeuGluLeuValAlaValGlyPheTyrGluProValLeuAlaSe 92
255 CTTCCCGCTTAGCAGAGTAGAGACACAGTTTCAGAGAGATAGGGAAGTTA 304
|||||:||||| ||| :|||:||||| |||:||||| |||
92 rIleProLysGluAspArgIleValGlnIleGluLysLeuLysGluPheA 109
305 AGAAGAGCTCTCGAGCTTCTCCCAAGGATTCTGGCTCCACAGAGCTC 354
|||||:||||| :||| |||:||||| |||:||||| |||
109 laArgAsnLeu...GlyTyrGluAlaArgGlyValTrpLeuThrGluArg 124
355 GCTATGACCGGATATCCCTGCCATCTAGGACACACGGTTATGAGTA 404
|||||:||||| ||| :|||:||||| |||:||||| |||
125 ValTrpGlnProGluLeuValLysSerLeuArgAlaAlaGlyIleAspTy 141
405 TCTATTCCGCGAGGGAGGCGATCTTTCTCAGCTCATCTCAACTCGG 454
|||||:||||| ||| :|||:||||| |||:||||| |||
141 rValIleValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGCCAATTAAACGGCTCTATCCACACCTTATAAGGCCCAA 501
|||:||||| ||| :|||:||||| |||:||||| |||
152 laGlyLeuSerLysAspGluLeuPheTyrProTyrTyrThrGluAspGly 168
502 AGGGA.....AAGCGCTTAGGTACATCAG 527
|||:||||| ||| :|||:||||| |||:||||| |||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArg..... 182
528 CTATCTCTTGGTCTCAGGAGCTTAGGAAGCGGATAAAGCTCGTTT. 576
|||||:||||| ||| :|||:||||| |||:||||| |||
183 .TyrLeuIleProPheArgProValAspLysThrLeuGluTyrLeuHis 199
577 .....GAAGTAGGTAAAGCTAAGGCGAGTC.....AAGACATC 612
|||||:||||| ||| :|||:||||| |||:||||| |||
199 erLeuAspAspGlyAspGluSerLysValAlaValPheHisAspAspGly 215
613 GAAGCGCTACCCGTTGGGTGGCGGTGAACACGGCTGTAATGCTCGGCAT 662
|||||:||||| ||| :|||:||||| |||:||||| |||
216 GluLysPheGlyValTrpProGlyThrTyrGluTrpValTyr..... 229
563 CGGAAGGCTTCTCTTATGATCTTAAGAAAGTGGCGAGCTGGATA.... 708
|||||:||||| ||| :|||:||||| |||:||||| |||
230 .....GluLysGlyTrpLeuArg 236
709 .....GAGGACAAGSACAACATCTTCTATAC 735
|||||:||||| ||| :|||:||||| |||:||||| |||

```

```

236 luPhePheAspArgValSerSerAspGluArgIleAsnLeuMetLeuTyr 252
736 GGCACCGGATATAGAG.....TTCAATTGGCTATAGGGACAT 770
|||:||||| ||| :|||:||||| |||:||||| |||
253 SerGluTyrLeuGlnArgPheArgProArgGlyLeuValTyrLeuProIle 269
771 TGCAGGCTAC...AGAATGAGCTGTGAGGGATTA..... 801
|||||:||||| ||| :|||:||||| |||:||||| |||
269 eAlaSerTyrPheGluMetSerGluTrpSerLeuProAlaArgGlnAlaL 286
802 .....TTAGAGGTTATACAGAGCTCAACTCGGAACCTGTCCTTCCC 843
|||:||||| ||| :|||:||||| |||:||||| |||
286 ysLeuPheValGluPheValGluGluLeuLys..... 296
844 TCAGAGCTGAAGCACAGTGAAGGAGGCTCTACTAGGACTTCGAGTTG 893
|||||:||||| ||| :|||:||||| |||:||||| |||
297 LysGluAsnLysPheAspArgTyrArgValPheValArgGlyGlyIleTr 313
894 GGCACAGATAAGAGCTTGAGGATATGGAGAGAGACGAAGGGAAC.... 939
|||:||||| ||| :|||:||||| |||:||||| |||
313 p.....LysAsnPhe...PhePheLysTyrProGluSerAsnTyrM 326
940 .....GCAAGACTTAATATGCTGCTTACATATGAGGGGCGAAGTCCG 984
|||||:||||| ||| :|||:||||| |||:||||| |||
326 eHisLysArgMetLeuMetValSerLysAlaValArgAsnAsn..... 340
985 CTTTATGCGGAGAACAGCGATGCAAGGGGATGGAGCCCTCCCTGAGAG 1034
|||||:||||| ||| :|||:||||| |||:||||| |||
341 .....ProGluAl 343
1035 GAGCGTGGATGCCCTCCGGCGGATATATAAGCAT.....TGGAGGGGTG 1078
|||:||||| ||| :|||:||||| |||:||||| |||
343 aArgGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGly 360
1079 AAAATGGG 1086
|||
360 alPheGly 362

seq_name: SwissProt_40:MALQ_THELI

seq_documentation_block:
ID MALQ_THELI STANDARD; PRT; 659 AA.
AC O32462;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase)
DE (Disproportionating enzyme) (D-enzyme).
GN JGT.
OS Thermococcus litoralis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=2365;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-19 AND 427-437, AND
RP CHARACTERIZATION.
RC STRAIN=DSM 5473;
RA MEDLINE=97454300; PubMed=9310375;
RX Jeon B.-S., Taguchi H., Sakai H., Ohshima T., Wakagi T., Matsuzawa H.;
RT "4-alpha-glucanotransferase from the hyperthermophilic archaeon
RT Thermococcus litoralis. Enzyme purification and characterization, and
RT gene cloning, sequencing and expression in Escherichia coli.";
RL Eur. J. Biochem. 248:171-178(1997).
CC -!- FUNCTION: Catalyzes the transglycosylation of
CC maltooligosaccharides, yielding maltooligosaccharides of various
CC lengths and glucose.
CC -!- CATALYTIC ACTIVITY: Transfers a segment of a (1,4)-alpha-D-glucan
CC to a new 4-position in an acceptor, which may be glucose or (1,4)-
CC alpha-D-glucan.
CC -!- ENZYME REGULATION: INHIBITED BY P-CHLOROMERCURIBENZOIC ACID,
CC MONIODOACETIC ACID, MERCURY AND NICKEL IONS.
CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
-----
CC

```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z72807; CAA97006.1; -  
 DR SGD: S0003255; MTL1.  
 KW Hypothetical protein; Transmembrane.  
 FT DOMAIN 104 107 POLY-SER.  
 FT DOMAIN 111 120 POLY-SER.  
 FT DOMAIN 124 138 POLY-SER.  
 FT DOMAIN 142 146 POLY-SER.  
 FT DOMAIN 182 188 POLY-SER.  
 FT DOMAIN 215 218 POLY-SER.  
 FT DOMAIN 222 232 POLY-SER.  
 FT DOMAIN 236 264 POLY-SER.  
 FT DOMAIN 269 275 POLY-SER.  
 FT DOMAIN 278 281 POLY-SER.  
 FT DOMAIN 284 291 POLY-SER.  
 FT TRANSMEM 362 382 POTENTIAL.  
 FT DOMAIN 468 474 POLY-ASP.  
 SQ SEQUENCE 551 AA; 57527 MW; 36D94341B319951 CRC64;

# alignment\_scores:

Quality: 115.50 Length: 290  
 Ratio: 0.791 Gaps: 10  
 Percent Similarity: 50.345 Percent Identity: 25.172

## alignment\_block:

US-09-886-400-3/rev x YGIF\_YEAST ..

Align seg 1/1 to: YGIF\_YEAST from: 1 to: 551

995 TCGGCTAAAGGCGGCTGCGGCTCATATTGTAGGACAGCATATTAAAG 946  
 ||||| ||||| ||||| ||||| |||||  
 69 SerSerAlaSerLeuSerThrProSerIleAla.....SerValSerPh 83  
 945 TCTTGGTTCCCTTCGTCCTCTCTCCATATCTCAAGCTCTTATCTGTGT 896  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 83 eThrSerPheProGlnSerSer...SerLeuLeuThrLeuSerThrL 99  
 895 CCGAATCGAAGCTAGTAGAGCTCCCTCCACTGTGCTTCAAGCTCT 846  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 99 euSerSerGluLeuSerSer...SerSerMetGlnValSerSerSer 114  
 845 GAGGGAAGGCACAGTTCGAGTGT.....AGCTCGTCTAT 811  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 115 ThrSerSerSerSerSerGluValThrSerSerSerSerSerSerI 131  
 810 AACCTCTATATCCCTCAACACTCATCTGTAGCTGCAATGCTCCCTAT 761  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 131 eSerProSerSerSerSerSerThrIle..... 141  
 760 AGCAATGAACCTATATCGGTGCGGTATAGAAGATGTTGCTGTGCC 711  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 142 .....SerSerSerSerLeuProThrPheThrValAlaSerThrSer 156  
 710 TCTATCCAGCTGCCACTTCTTAGGATCATAGAGAACCTTCCGAT 661  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 157 SerThrValAlaSerSerThrLeuSer...ThrSerSerSerLeuVal 172  
 660 GCGGACATACAGCGGTCTTACGGCCACCAACAGGCTACGCTTCA 611  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 172 eSerThrSerSerSerThrPheSerSerSerGluSerSerSerSerL 189  
 610 TGTCTTGAAGCTTACGTTAGCTTACCTTCAAAACAGAGCTTTATC 561  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 189 euLeuSerSerIleSerThrSerValSerThrSerSerValThrVal 205

560 GCCTTCCTAAGCTCCCTGAGACCAAGAGATAGCTGATGTACCTAAAGCG 511  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 206 ProSerSerSerThrSerSerProPro..... 214  
 510 CTTTTCCTTTGGGCTTTATAAGGTGTGATAGAGCGGTTTAAATGGCT 461  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 215 .....SerS 216  
 460 TTATCGCGGAGTTGAGATGAGTGAAGAAAGCATCGCTCCCGTCGGCG 411  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 216 erSerSerGluLeuThrSerSerSerTyrSerSerSerSerSerSer 232  
 410 AATAGATACTATACACCGTTGCTTCAGTATGCGAGGATATATCGGGTC 361  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 233 ThrLeuPheSerTyrSerSerPheSer..... 242  
 360 ATAGCGGAGCTGCGCAGCCAGATCCCTTTGGAGAAAGCTCGAAGAGCT 311  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 243 ...SerSerSerSerSerSerSerSerSerSerSerSerSerS 258  
 310 CTTCTTAACTTCCCTATCTCTGAACTTGTGCTTACTCTCTGCTAAGC 261  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 258 erSerSerSerSerSerTyr.....PheThrLeuSerThr 270  
 260 GGGAGAGGGGAGTATTCGTCGCTGTAGCTCGTTCGGATATCTCTAT 211  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 271 SerSerSerSerSerIleTyrSerSerSerSerTyrProSerPheSer 287  
 210 CAGGTCACTCGCGATCGCCCTTTAACGAGGTCTATATATATCTCTCGGA 161  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 287 rSerSerSerSerAsnPro.....ThrSerSerIleThrSerThrSerA 302  
 160 GGAACCTTAAAGTATAGCCC 141  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 302 laSerSerSerIleThrPro 308  
 seq\_name: SwissProt\_40:MDH\_BACSU  
 seq\_documentation\_block:  
 ID MDH\_BACSU STANDARD; PRT; 311 AA.  
 AC P49814;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Malate dehydrogenase (EC 1.1.1.37) (Vegetative protein '69) (VEG69).  
 GN CITH OR MDH.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / SMV;  
 RX MEDLINE=95134995; PubMed=8550482;  
 RA Jin S., de Jesus-Berrios M., Sonenshein A.L.;  
 RT "A Bacillus subtilis malate dehydrogenase gene."  
 RL J. Bacteriol. 178:560-563(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98048467; PubMed=9387221;  
 RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;  
 RT "Sequencing and functional annotation of the Bacillus subtilis genes  
 in the 200 kb rnaB-dnaB region."  
 RL Microbiology 143:3431-3441(1997).  
 RN [3]  
 RP SEQUENCE OF 1-15.  
 RC STRAIN=IS58;  
 RX MEDLINE=97443988; PubMed=9298659;  
 RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,  
 RA Hecker M.;  
 RT "First steps from a two-dimensional protein index towards a response-  
 regulation map for Bacillus subtilis."  
 RT Electrophoresis 18:1451-1463(1997).  
 RL



```

30 laaLaSerThrSerLeuSerSerSerVal..... 40
686 GGATTCAAGAGGAGCGTCCGATCGGACATTACAGCGGTTCAC 637
41 .....ilPrSerSerSerSerMetLeuSerSe 51
636 GCCACCCAAACGGGTAGCGTTCGATGTC.....TTGACTG 599
51 rSerSerAlaThrAlaLeuSerSerSerSerSerProLeuSerS 68
598 CTTTACGTTACCTTACCTTCAAAAGGACCTTATCGCCTTCCTAAGC 549
68 exSerSerPheThrSerProAlaSerSerPheLeuThrSerLeuVal 84
548 TCCCTGACACCAAGGAGATAGCTGATGATACCTAAAGGCGCTTTCCTTGG 499
85 SerSerSerSerGln.....GlnSerSerSerSe 95
498 GGCCTTTATAGGTGTGATAGCGGTTTAATTTGGCTTTATCCCGAGT 449
95 rAlaSerLeuThrSerSerSerSerAlaThrLeu..... 106
448 TGAGATGAGCTGAGAAAGCATCGCTCCCGTCGGGGAATAGATACTCA 399
107 .....ThrSerSerSerSerAlaSerProThrSerSerSerSer 120
398 TAACCGTTGCTTCAGTATGCGAGGATTCGCGTATAGCGGAGCTC 349
121 HisAlaLeuSerSerSerSerSerLeuValAlaSerSerSerSe 137
348 TGGC.....AGCCAGAATCCCTTTGGAG 326
137 rGlyMetSerSerSerSerLeuSerHisSerSerSerValProSerSers 154
325 AAAGCTG...AAGAGCTTCCTTAACCTCCCATCTCTGTGAACTGT 279
154 exSerSerThrHisSerSerSerMetThrThrSerGlyLeuSerSer 170
278 GCTTCTACTCTGCTAAGCGGAGGAGGGAGTATCGCTCGGTAGCT 229
171 AlaSerIleValSerSerThrTyrArgAspGlyProSerIleIleThrLe 187
228 CGTTCGGATTATCTATCAGGTCTACTCGCATGCCCTTAACTTACG... 183
187 uValSerThrSerTyrValSerGluValValThrProThrThrAsnA 204
182 .....AGGTCTAATATCTTCGGGAGAACTTAAAGTATAGCCGCTT 138
204 snTrpAsnSerSerSerSerPheThrSerSerThrSerThrProile 220
137 ATGTTGAGCCCAAGAAATTCCTTTAATCAGTCTCGATGCTGACTGG 88
221 SerSerSerTyrSerSerSerGlyThrLeuProSerSerSerSerLysSe 237
87 GATGATGCCCTTCTATGACCTTTGGATTTCG..... 54
237 rSerAsnHisValGlyValValValGlyCysSerValAlaIleProValG 254
53 .....CTCTTTGGATTTCGGCATAC...TGGAGG 27
254 llyValValLeuLeuLeuGlyLeuGlyIlePheLeuTrpLys 268
seq_name: SwissProt_40:MDH_BACHD
seq_documentation_block:
ID MDH_BACHD STANDARD; PRT; 314 AA.
AC Q9K849;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH OR CITH OR BH3158.

```

```

OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxaloacetate + NADH.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP001517; BAB06877.1; -
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR001252; MDH_actsite.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001236; ldh.
DR Pfam; PF002856; ldh; 1.
DR PRINTS; PR00086; LLDHGRNASE.
DR PROSITE; PS00086; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.
FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 314 AA; 33680 MW; 02D132F11B3B8E34 CRC64;
alignment_scores:
Quality: 103.50 Length: 217
Ratio: 1.005 Gaps: 9
Percent Similarity: 47.465 Percent Identity: 23.041
alignment_block:
US-09-886-400-3 x MDH_BACHD ..
Align seg 1/1 to: MDH_BACHD from: 1 to: 314
28 CTCACGATGCGGAATCCCAAGAGGGAATCCCAAG..... 66
130 MetThrTyrThrValTyrLysGlySerGlyPheProLysAsnArgValI 146
67 .....GTCAATAGAGAGGCATACATCCAGTCATCGAGACAC 103
146 eclyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
104 TGATTAAAGAGAAATTCCTTTGGGCTC...AACATAACGGGCTATACC 150
160 heValAlaGlnGluLeuAsnLeuSerValGluAspIleThrGlyPheVal 176
151 TTAAGATTCTCCCGAAGGATATTATAGACTCGTTAAAGGGGCGCATCG 200
177 Leu.....GlyGlyHisG 181
201 GAGTCACCTGATAGAGATAATCGGAACGACGATACACGACGACATA.... 246
181 yAspAspMetValProLeuIleA-gTyrSerTyrAlaGlyIleProL 198
247 ..CTCCCGCTCCTCCGCTTAGCAGAGTAGAAGCAACAAAGTTCAGAGAGAT 294

```

```

198 euGluLysLeuLeuProGlnGluArgIleAspAlaIleValGluArgThr 214
    ||||| ||||| :|||:||||| ||||| |||||
295 AGGAAGTTAAGGAAGAGCTCTCGAGCTTTCCAAAGGGA.....TT 338
    ||||| ||||| ||||| ||||| ||||| |||||
215 ArgLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 231
    ||||| ||||| ||||| ||||| ||||| |||||
339 CTGGCTGCCAGAGCTCGCTATGACCCGATAATCCCTGCCATPACTGAAG 388
    ||||| ||||| ||||| ||||| ||||| |||||
231 rTyAlaProAlaAlaSerLeuAlaGluMetValGluAlaIleLeuLys 248
    ||||| ||||| ||||| ||||| ||||| |||||
389 ACAACGGTTATGAGTATCTATCTGCGACGGGAGCGGATGCTTTCTCA 438
    ||||| ||||| ||||| ||||| ||||| |||||
248 sPlys..... 249
439 GCTCATCTCACTCGCGGATAAGCAATTAACCGCTCTATCCACACT 488
    ||||| ||||| ||||| ||||| ||||| |||||
250 .....LysArgValLeuProThrIleAlaTyrLe 259
489 TATAAGGCCCAAGGAAGCGCTTTAGGTAC..... 522
259 u.....GluGlyGluTyrGlyTyrGluAspIleTyrValG 271
523 .ATCAGCTATCTCTGTGCTCAGGAGCTTAGGAAGCGGATAAGCTC 570
    ||||| ||||| ||||| ||||| ||||| |||||
271 lyValProThrIleLeuGlyGlyAspGlyIleGluLysValIleGluLeu 287
571 GTTTTCAGGTAGGTACGCTTAAGGCGAGTCAGACATGGAAGCGGT 620
    ||||| ||||| ||||| ||||| ||||| |||||
288 AspLeuThrAspGluGluLysAlaThrPheAlaLysSerIleGluSerVa 304
621 A 621
304 1 304

```

seq\_name: SwissProt\_40:YNR6\_YEAST

```

seq_documentation_block:
ID YNR6_YEAST STANDARD; PRT; 636 AA.
AC P53882;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 67.4 kDa protein in RPS3-PSD1 intergenic region.
GN YNL176C OR N1661.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M., Dondey H.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 271452; CAA96068.1; -.
DR SGD: S0005120; YNL176C.
DR Hypothetical protein; Transmembrane.
KW DOMAIN 94 98 POLY-SER.
FT DOMAIN 131 135 POLY-SER.
FT DOMAIN 301 309 POLY-SER.
FT DOMAIN 317 322 POLY-SER.
FT TRANSMEM 457 477 POTENTIAL.
SQ SEQUENCE 636 AA; 67356 MW; 4027EF48DEA5F2B7 CRC64;

```

```

alignment_scores:
  Quality: 103.50      Length: 343
  Ratio: 0.681        Gaps: 16
  Percent Similarity: 44.315      Percent Identity: 23.324

alignment_block:
US-09-886-400-3/rev x YNR6_YEAST ..
Align seg 1/1 to: YNR6_YEAST from: 1 to: 636

1088 TCCCCATTTTCCACCTCCCAATCGTATATATATCGCCGGAAGGCATCCAG 1039
    ||| :|||:||||| ||||| |||||
105 SerGluTyrAspProAlaThrSerSerTyr.....Se 115
1038 CTCCTCTCTCAGGAGGGCTCCCATCCCTTCGATCGCTGTCTCGGCTA 989
    |||||:||||| ||||| |||||
115 rIleIleThr.....ProSerMetSerIlePheSer.... 125
988 AAAGGCGGAGTTTCGCCCTCATATTGTAGGACAGCATATTAGTCTTCG 939
    |||||:||||| ||||| |||||
126 .....SerThrSerProMet.....SerSerSerSerIleThr 137
938 TTCCTCTCTCTCTCTCCATATCTCAAGCTCTTATCTGTGGTGCCTAAC 889
    |||||:||||| ||||| |||||
138 SerGluTyrSerSerLeuThr..... 144
888 CGAAGTCCGTAGTAGAGCTCCCTCCACTGTCTCAGCTCTGAGGAA 839
    |||||:||||| ||||| |||||
145 .....SerThrProThrLeuSerSerSer..... 153
838 GGCACAGTTCGAGTTCAGCTCGTCTATAACCTCT...AATAATCCCTCA 792
    |||||:||||| ||||| |||||
154 .....AlaThrSerLeuSerSerSerTyrSerSerLeuSerSerProSer 168
791 ACATCATCTCTGTAGCTGCAATGTCCTATAGCCATAGACTCTATATC 742
    |||||:||||| ||||| |||||
169 SerLeuLeuValSerSerLeuSerLeu.....SerLeuSe 181
741 GGTCCGCTATAGAAGAATCTTCTCTCTATCCAGCTCGCCACTT 692
    ||| |||||:||||| ||||| |||||
181 rSerSerTyr.....SerAspThrLysLeuPheSerP 192
691 TCITAGATTCATAAGGAAGCTTCGATCGGAGCATTTACAGCCGTG 642
    ||| |||||:||||| ||||| |||||
192 heAspSer.....ArgSerSerIlePheSerProSerThrProThrVal 206
641 TTCAGGCCACCAAGCGGTACGCTTCGATGTCTTTGACGCTTTAG 592
    |||||:||||| ||||| |||||
207 IleSerProSerTyrThrTyrLeuSerSerIleSerAlaThrSerPheG 223
591 CGTTACCTTACCTTCAAAACGAGCTTTATCGCCTTCCTAAGCTCCCTGA 542
    |||||:||||| ||||| |||||
223 nIleSerThrSerGluLeuSerSerSerSerTrpPheSerThrIleSerS 240
541 GACCAAGGAGATAGCTGATGTACCTAAAGCGCTTTTCCCTTTGGGCTTT 492
    ||| |||||:||||| ||||| |||||
240 erProSer..... 242
491 ATAAGTGTGGATAGACGGCTTTAATGCTTTATCGCCGAGTTGAGATG 442
243 .....ThrIl 244
441 AGCTGAGAAAAGCATCGCTCCCGCTCGCGGAATAGATACTCATACCGT 392
    |||||:||||| ||||| |||||
244 eSerAsnLysAspThrThrCysProSerSerSerArgAsnThrSerThrS 261
391 TGTCTCTTCAGTATGGCAGGATATCGGCTCATAGCGAGCTCTGGCAGC 342
    |||||:||||| ||||| |||||
261 erPheTyrSer.....SerSerLeuSerSer 269
341 CAGATATCCCTTTGGAGAACGTCGAGAGCTCTTCTCTTAACCTTCCCTATC 292
    |||||:||||| ||||| |||||
270 ThrAsnAspPheSerThrIleSerLysSerLysLeuSerProSerAl 286

```



```

alignment_scores:
  Quality: 101.50      Length: 465
  Ratio: 0.543        Gaps: 22
  Percent Similarity: 40.215  Percent Identity: 16.559

alignment_block:
  US-09-886-400-3 x RNR_AQUAE  ..

  Align seg 1/1 to: RNR_AQUAE from: 1 to: 705

13  GTCTTTTCCACGGCACTCCAGTATGCCGAATCCCA.....AA 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
100 ValPheHisGlyAspValValLysAlaLysValThrGluPheLysGly 116
   51 GAGCGAA.....ATCCCAAGTCATAGAGAAGGCATACATCCAGTCA 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
116 sLysGluValArgIleIleLysValLeuLysArgAlaLysLysAspIleV 133
   95 TCGAGACACTGATTAAAGAAAGAA.....ATTCCTTTTGGG 129
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
133 alaLalLysValValPheGluAspGluGlnCysTyrValValProLeuasp 149
   130 CTCACATAACGGCTATACCTTAAAGTTCCTCCCGAAGATATTATAGA 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
150 GluAsn...AlaHisHisArgIleLeuLeuSerLysLysAspCysGlnly 165
   180 CCTCGTTAAAGGGGGCATCGCG..... 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
165 sLeuLysGluGlyGluValValValLeuLysIleThrGlnPheProThrL 182
   202 .....AGTGACTGATAGATAATCGGA..... 225
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
182 yLysSerProAlaArgGlyLysValIleGluValLeuGlyAsnProLys 198
   226 .....AC 227
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
199 GluLysPheIleAlaIleAspValIleIleArgLysTyrAsnLeuProth 215
   228 GAGCTACAGCAGCCAAATACTC.....CCCTCTCCCGCGTATGACA 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
215 rSerTyProGluLysValIleLysGluValGluAlaIleProGluGluI 232
   269 GAGTAGAAGCACAAAGTTCAGAGAGATAGGAAGTTAAGGAAGAGCTC... 315
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
232 leProGluGluGluIleLysArgArgAspLeuArgGluGlnLeuCys 248
   316 .....TT 317
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
249 PheThrIleAspProGluLysAlaGlyAspPheAspAlaValAlaIle 265
   318 CGAGCTTCTCCAAAGGATCTCTGG.....CTGCCAGAC 352
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
265 eGluLeuThrProGluGlyTyrTyrLysLeuTyrValHisIleAlaAspV 282
   353 TCGCCTATGACCCGATAATCCCTGCCATACCTGAAGCAC..... 390
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
282 alSerTyr.....TyrValArgGluGlyThrGluThr 292
   391 .....AAGCTTTATGATCTATCTATCCGACGGGGA 422
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
293 AspLysGluAlaTyLysArgGlyPheThrTyrTyProAspArgAl 309
   423 GGCATGCTTTTCTACGCTCATCTCACTCGGCGATAAAGCCNATTAAAC 472
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
309 aLeuHisMetLeuProGluLysLeuSerAlaLysLeuCysSerLeuArgP 326
   473 CGCTCTATCCACACTTATAAGGCCCAAGGAAAGCGCTTTTAGGTAC 522
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
326 ro.....AsnGluAspLysLeuAlaPhe 333
   523 ATCAGCTATCTCTTGCTCAGGAGCTTTAGGAAGCGGATAAAGCTCGT 572
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
334 .....ThrValGluMetVa 338

```

```

seq_documentation_block:
ID  ID  RNR_AQUAE  STANDARD;  PRT;  705 AA.
AC  O67834;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Ribonuclease R (EC 3.1.-.-) (RNase R) (VaccB protein homolog).
GN  RNR OR VACB OR AQ_2046.
GE  Aquifex aeolicus.
OC  Bacteria; Aquificales; Aquificaceae; Aquifex.
OX  NCBI_TaxID=63363;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=VF5;
RX  MEDLINE=98196666; PubMed=9537320;
RA  Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA  Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus.";
RL  Nature 392:353-358(1998).
CC  -!- FUNCTION: 3'-5'EXORIBONUCLEASE THAT PARTICIPATES IN AN ESSENTIAL
CC  CELL FUNCTION. ACTS NONSPECIFICALLY ON POLY(A), POLY(U) AND
CC  RIBOSOMAL RNAs (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.
CC  -!- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AE000769; AAC07792.1; .
CC  InterPro: IPR002059; Cold_shock.
CC  InterPro: IPR001500; Ribonuclease_II.
CC  InterPro: IPR003029; SI.
CC  DR  Pfam; PF00773; RNB; 1.
CC  DR  Pfam; PF00575; SI; 1.
CC  DR  SMART; SM00357; CSP; 1.
CC  DR  SMART; SM00316; SI; 2.
CC  DR  PROSITE; PS01175; RIBONUCLEASE_II; 1.
CC  DR  PROSITE; PS0126; SI; 1.
CC  KW  Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.
CC  DOMAIN 615 696 SI MOTIF.
CC  FT  SEQUENCE 705 AA; 81537 MW; 2C08587229D1B05E CRC64;
SQ

```

```
573 TTTTCAAGGTAAGCTAACGCTAAAGCAGCTCAAGACATCGAGCGCTA. 621
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 lPheAspGluSerGlyAsnLeuLysAlaTyrAspIleTyrGluSerVali 355
622 .....CCGGTTGGTGGCGTGAACACGCGCTAATGCTCGCATC 663
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 leArgSerLysAlaArgLeuThrTyrAsnGluAlaLeuAlaLeuVal 371
664 GGA.....AGGCTCTCT...CTTATGAATCTTAAGAA 692
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 GlyAspProAlaLeuGluLysLysPheProAsnLeuValGluProLeuAr 388
693 AGTGGCGGCG. ....TGS ..... 705
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 gMetMetGluThrLeuTyrArgIleLeuSerArgLysArgTyrGluMetG 405
705 ..... 705
405 lySerIleAspPheAspLeuProGluAlaGluValIleValAspGluTyr 421
705 ..... 705
422 GlyGluProThrAlaIleTyrProTyrGluArgHisValAlaHisArgI 438
706 .....ATAGAGGACAGGACACATCTTCTATAGGCA 739
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 eIleGluHisPheMetIleSerAlaAsnGluThrValAlaLeuHis.... 453
740 CCGATATAGATTCATGCTATAGGACATTCAGGCTACAGATG... 786
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454 .....LeuGluHisAlaGlyTyrProCysLeu.....TyrArgValHis 466
787 .....AGTGTGAGGATTATTAGAGTTATAGACCA 818
467 GluProProAspGluGluLysValGluAsnLeuLeuGluIleLeuGlu 483
819 GCTCACTCGGACTGCGCTTCCCTCAGAGCTGAAGCAGTGAAGGG 868
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 yLeuGlyTyrLysValLysArgProHisGlu..... 493
859 AGCTCTACTTACGGACTCGAGTGGCCACAGATAAGAGCTTGAGGATA 918
494 .....TyrThrProLysPhePheGlnLysIle 502
919 TGGAGAGAGCAGCAGGAGGACGCAACTTAATATGCTGCTCTAC 963
||||: ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 lIleGluAspPheGluGlyArgProGluGluAsnLeuValArgPhe 517

seq_name: SwissProt_40: LONM_SCHPO
seq_documentation_block:
ID LONM_SCHPO STANDARD; PRT; 1067 AA.
AC Q09769;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ATP-dependent protease, mitochondrial precursor
DE (EC 3.4.21.-).
OS SPAC22F3.06C.
ON Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lye G., Churcher C.M., Barrell B.G., Rastandream M.A., Walsh S.V.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases
CC -!- FUNCTION: REQUIRED FOR INTRAMITOCHONDRIAL PROTEOLYSIS. CATALYZES
CC THE INITIAL STEPS OF PROTEIN DEGRADATION (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
```

```
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE
CC LON FAMILY OF ATP-DEPENDENT PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z54285; CAA91071.1; ..
CC DR MEROPS: S16.002; ..
CC DR InterPro: IPR003593; AAA.
CC DR InterPro: IPR003599; AAA_subfam.
CC DR InterPro: IPR003111; LON.
CC DR InterPro: IPR001984; Lon_endopep.
CC DR Pfam: PF00004; AAA; 1.
CC DR Pfam: PF02190; LON; 1.
CC DR PRINTS: PR00830; ENDOLAPTASE.
CC DR SMART: SM00382; AAA; 1.
CC DR SMART: SM00464; LON; 1.
CC DR PROSITE: PS01046; LON_SER; 1.
CC KW Hypothetical protein; Hydrolase; Serine protease; ATP-binding;
CC Mitochondrion; Transit peptide.
CC FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
CC FT CHAIN 1 1067 PUTATIVE ATP-DEPENDENT PROTEASE.
CC FT NP_BIND 578 585 ATP (POTENTIAL).
CC FT ACT_SITE 946 946 BY SIMILARITY.
CC FT SEQUENCE 1067 AA; 118641 MW; 215FCFBE9C0C4AD CRC64;
CC SQ

alignment_scores:
Quality: 101.00 Length: 454
Ratio: 0.561 Gaps: 22
Percent Similarity: 39.648 Percent Identity: 21.145

alignment_block:
US-09-886-400-3 x LONM_SCHPO ..
Align seg 1/1 to: LONM_SCHPO from: 1 to: 1067

53 AAAGGTCATAGAGAGGCATACATC.....CCAGTCATCG 97
614 LysGlyHisArg.ArgThrTyrIleGlyAlaMetProGlyLysIleValG 630
98 AGACACTG.....ATTAAAGAAATTCCTTTGGCTCAACATAAGC 141
630 lAlaLeuLysLysValGlnThrGluAsnProLeuIleLeu..... 643
142 GGCTATACCTTAAGTCTCCCGAAGGATATATAGACCTCGTT..... 186
644 .....IleAspGluIleAspLysValGly 652
187 .....AAAGGGGCATCGCAGTACCTGATAGAGATAATCGGAACGA 229
652 sSerHisGlnGlyAspProAlaSerAlaLeuGluLeuLeuAspSerG 669
230 GCTACACGACGCAACTACTCCCTCC.....CTCCCGCTTAGCAGA 270
669 luGlnAsnSerAlaPheLeuAspTyrTyrMetAspIleProLeu..... 683
271 GTAGAAGCACAAAGTTCAGAGAGATAGGAGGTAAAGAGAGCTCTCGA 320
684 .....AspValSerValLeuPheVa 691
321 GCTTCTCCAAAGGATCTGCTGGTCGACAGCTCGCTATGACCCGATAA 370
691 lCysThrAlaAsn.....ThrIleAspThrIleP 701
371 TCCCTGCCACTACG.....AAGGACACAGCTTAT... 399
701 roProProLeuLeuAspArgMetGluValIleGluLeuSerGlyTyrVal 717
```

```

400 .....GAGTATCTATTGCGCGACGGGA 422
718 SerAlaGluLysValAsnIleAlaLysGlyTyrLeuIleProGlnAlaLys 734
423 GCGGATG.....CTTTCTCAGCTCATCTCACTCGCGCATAAAGCCAA 466
734 salaalacysGlyLeuLysAspAlaAsnValAsnIleSerAspAlaI 751
467 TTAACCCGCTCTATCCACACCTTATAAGGCCCAAGGAAAGCGCTTT 516
751 leuLysGlyLeu.....Ile 755
517 AGGTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAAGCGCATAAA 566
756 SerTyrTyrAlaHisCysGlyValArgAsnLeuLysLysSerIleG 772
567 GCTCGTTTGAAGTAAGTAAGCTAAAGGCGAGCTCAAGACATCGAAG 616
772 uLysIlePheArg.....LysThrSerPheSerIleValLysGluIleAsp 788
617 CGGTACCCGCTTGGTGGCGCGTGAACAG..... 645
788 sp.....GluLeuAsnSerLysGluLysSerThrGlyLys 799
646 .....GCTGTAATGCTCGG 659
800 SerGlyLysLysThrSerProGlnSerSerGluAspAlaAlaAsnLysG 816
660 CATCGGAGGCTCTCTTATGAATCCCAAGAAAGTCGCGAGCTGGATG 709
816 uAlaSerSerValProLeuLysValProAspLysValAsnIleGluIleG 833
710 AGCAACAAGGAC.....ACATT 726
833 luGluLysAspLeuThrLysTyrLeuGlyProIleTyrThrSerGln 849
727 CTTCTATACGCGACCGAT.....ATAGATTCTATGGCTATAG 764
850 ArgLeuTyrAspThrThrProGlyValValMetGlyLeuGlyTyrTh 866
765 GGACATTCGAGCTACAGAATGAGTGTTCAGGAGTATTAGAGCTTATAG 814
866 rProMetGlyGlyValSerMetTyrValClnThrIleValLysAsnIleL 883
815 AGGAGCTCAACTCGGAACGTGCTCTT..... 840
883 euSerSerAsnSerThrProSerLeuGluArgThrGlyClnLeuGlyAsp 899
841 .....CCCTCAGCTGAGCACAGTGGAGGAGCTCTACTT 878
900 ValMetLysGluSerSerGluIleSerTyrSerPheSerLysSerPheLe 916
879 A.....CGGACTTCGA 889
916 uSerLysHisPheProAsnAsnLysPhePheClnHisAlaArgLeuHisM 933
890 GTTGGCCACCATAGAGCTTTGAGGATATGAGAGAGGACGAGGGAAC 939
933 eHisCysProGluGlySerIle.....SerLysAspGlyProSer 946
940 GCAAGACTTAATG..... 954
947 AlaGlyIleThrMetAlaThrSerLeuLeuSerLeuAlaLeuAspThrPr 963
955 .....CTGTCTCAATATGAGGGGGAACCTCGCCCTT..... 987
963 oValProAlaThrThrAlaMetThrGlyGluLeuThrLeuThrGlyLysI 980
987 ..... 987
980 leuLeuArgIleGlyLeuArgGluLysThrValAlaAlaLysLeuSer 996

```

```

988 .....TTAGCGGAGAACGCCATGCAAGGGGATGGGAGCC 1022
997 GlyMetLysGluIleLeuPheProLysSerAsnLeuAlaAspTrpGluG 1013
1023 CCTCCCTGAG 1032
1013 nLeuProAsp 1016
seq_name: SwissProt_40:MDH_BACTC
seq_documentation_block:
ID MDH_BACTC STANDARD; PRT; 312 AA.
AC Q9X4K6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH.
OS Bacillus thermodenitrificans.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.
OX NCBI_TaxID=33940;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=00462;
RA Williams R.A.D., Welch S.G., Alawadhi S.A.;
RT "Properties and primary structure of a thermostable L-malate
RL dehydrogenase from 'Bacillus thermodenitrificans'";
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) -> oxaloacetate + NADH.
CC -1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF114423; AAD28555.1;
CC HSPF; Q27743; ICET.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR001252; MDH_actsite.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001236; ldh.
DR Pfam; PF000056; ldh; 1.
DR Pfam; PF02866; ldh_C; 1.
DR PRINTS; PR00086; LLDHDEGNASE.
DR PROSITE; PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
KW ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33682 MW; 68C691C462EFF452 CRC64;
alignment_scores:
Quality: 100.50 Length: 216
Ratio: 0.966 Gaps: 9
Percent Similarity: 48.148 Percent Identity: 22.222
alignment_block:
US-09-886-400-3 x MDH_BACTC ..
Align seg 1/1 to: MDH_BACTC from: 1 to: 312
28 CTCGAGTATGCGGAAATCCAAAGAGCGAAATCCCAAG..... 66
130 MetSerTyrThrValPheLysGluSerGlyPheProLysAsnArgValI 146
67 .....GTCAATAGAGCGCATACATCCCATCGTATCGAGAC 103

```

```
146 eGlyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
104 TGATTAAGAAGAAATTCCTTTGGGCTCAACATAACGGGCTATACCTTA 153
    : : : : : : : : : : : : : : : : : : : : : : : :
160 heValAlaGlnGlu.....LeuAsnIleSer..... 168
154 AAGTTCCTCCGGAAGGATATATAGACCTCGTTAAAGGGGCATCCGAG 203
    : : : : : : : : : : : : : : : : : : : : : : : :
169 .....ValLysAspValThrGlyPheValLeuGlyHisGlyAs 182
204 TGACCTGATAGATATATCGGAACGACGTACACGCACGCAATA.....C 247
    : : : : : : : : : : : : : : : : : : : : : : : :
182 pAspMetValProLeuValArgTyrSerTyrAlaGlyIleProLeuG 199
248 TCCCCCTCCCTCCGCTTAGCAGAGTAGAGCACAAAGTTCAGAGATAGG 297
    : : : : : : : : : : : : : : : : : : : : : : : :
199 luLysLeuIleProLysAspArgLeuAspAlaIleValGluArgThr 215
298 GAAGTTAAGGAAGAGCTCTTCGAGCTTTCCTCAAAGGGA.....TTCG 341
    : : : : : : : : : : : : : : : : : : : : : : : :
216 LysGlyGlyGlyGluIleValasnLeuLeuGlyAsnGlySerAlaTyr 232
342 GCTCCAGAGCTCCCTATGACCCGATATCCCTGCCATCTACTGAGGACA 391
    : : : : : : : : : : : : : : : : : : : : : : : :
232 rAlaProAlaAlaSerLeuAlaGluMetValGluAlaIleValLys 249
392 ACGCTTATGATCTATTCCCGACGGGAGCGGATGCTTTTCTCAGCT 441
    : : : : : : : : : : : : : : : : : : : : : : : :
249 in..... 249
442 CATCTCAACTCGCGATAAAGCCAATTAACCGCTCTATCCACACCTTAT 491
    : : : : : : : : : : : : : : : : : : : : : : : :
250 .....ArgArgIleLeuProAlaIleThrTyrLeu... 259
492 AAAGGCCCAAGGAAAGCCCTTAGGTAC.....A 523
    : : : : : : : : : : : : : : : : : : : : : : : :
260 .....GluGlyGluTyrGlyTyrGluGlyIleTyrLeuGlyV 272
524 TCAGCTATCTCCTGGTCTCAGGAGCTTAGGAAGCGCATAAAGCTCGTT 573
    : : : : : : : : : : : : : : : : : : : : : : : :
272 alProThrIleLeuGlyAsnGlyIleGluLysValIleGluLeuGlu 288
574 TTGAGGTAAAGTAAGCTAAAGGCGAGTCAAGACATCGAAGCCGTA 621
    : : : : : : : : : : : : : : : : : : : : : : : :
289 LeuThrGluAspGluLysAlaAlaLeuAlaLysSerLeuGluSerVal 304
```

OM of: US-09-886-400-3 to: Issued\_Patents\_AA:\* out\_format : pfs  
Date: Jun 11, 2002 10:34 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=framet-n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_Spool/US09886400/runat\_11062002\_150914\_17298/app\_query.fasta\_1.1169  
-DB=Issued\_Patents\_AA -QFMT=fastcan -SUFFIX=ra1 -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09886400 -CGNJ\_1\_31 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-886-400-3  
Query length: 1095  
Database: Issued\_Patents\_AA:\*  
Database sequences: 231628  
Database length: 2442594  
Search time (sec): 25.060000

score_list:	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-613-220B-4 +	1580.00	3187.30	1.4e-170	346	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-894-212A-8 +	141.50	250.84	2.7e-07	647	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-894-212A-2 +	141.50	250.81	2.7e-07	649	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-893-938A-1 +	141.50	250.79	2.7e-07	650	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-270-013B-2 +	104.00	182.69	0.0033	329	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-838-418-2 +	104.00	182.69	0.0033	329	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-228-986-72 -	93.00	147.96	0.1030	910	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-061-709-2 -	92.50	144.19	0.1331	1142	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-936-135-4 +	91.50	132.27	0.2717	2584	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-418-893D-23 +	88.50	128.73	0.5286	2089	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-418-893D-24 +	88.50	128.73	0.5286	2089	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-355-166-1 +	87.50	140.71	0.3617	657	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-141-135-2 +	87.00	152.15	0.2332	235	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-248-021A-2 +	86.50	138.10	0.4822	639	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-910-923-3 +	86.00	136.17	0.5729	743	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-235-451-36 +	85.50	134.81	0.6629	784	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-907-166-6 +	85.50	134.44	0.6744	788	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-993-118-7 -	84.00	126.89	1.23	1142	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-845-528C-7 -	84.00	126.89	1.23	1142	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-07-718-533-3 +	83.50	132.87	1.01	641	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-161-999-3 +	83.50	132.87	1.01	641	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-487-28A-2 +	83.50	121.22	1.73	1676	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-849-480A-5 +	83.00	134.71	1.01	506	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-564-803-237 +	83.00	131.94	1.15	636	
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-08-995-280C-2 -	82.50	140.52	0.8454	288	
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-09-215-042-2 -	82.50	140.52	0.8454	288	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-676-279-59 +	82.50	132.71	1.21	549	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-804-871-1 +	82.50	132.60	1.21	554	
/cgn2_6/ptodata/1/iaa/6A_COMB.ppt:US-09-002-298-1 +	82.00	138.89	0.9508	303	
/cgn2_6/ptodata/1/iaa/6B_COMB.ppt:US-09-104-068-4 +	82.00	130.16	1.48	623	
/cgn2_6/ptodata/1/iaa/6B_COMB.ppt:US-09-104-068-2 +	82.00	129.89	1.50	637	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-489-733-3 +	81.50	132.02	1.47	491	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-993-581B-3 +	81.50	132.02	1.47	491	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-451-715A-6 +	81.50	126.02	1.94	806	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-09-412-545-2 +	81.00	121.62	2.58	1065	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-484-493-2 +	80.50	128.61	2.04	550	
/cgn2_6/ptodata/1/iaa/5A_COMB.ppt:US-08-484-494-2 +	80.50	128.61	2.04	550	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-345-212-2 +	80.50	128.61	2.04	550	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-09-249-003-2 +	80.50	128.61	2.04	550	
/cgn2_6/ptodata/1/iaa/5B_COMB.ppt:US-08-468-558-2 +	80.00	122.25	2.97	855	

/cgn2\_6/ptodata/1/iaa/5B\_COMB.ppt:US-08-676-444-2 + 80.00 122.25 2.97 855  
/cgn2\_6/ptodata/1/iaa/5A\_COMB.ppt:US-08-118-441-29 + 80.00 119.09 3.43 1110  
/cgn2\_6/ptodata/1/iaa/6A\_COMB.ppt:US-08-338-579A-29 + 80.00 119.09 3.43 111  
/cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.ppt:PCT-US94-09851-29 + 80.00 119.09 3.43 831  
/cgn2\_6/ptodata/1/iaa/5A\_COMB.ppt:US-08-073-384C-5 + 79.50 121.58 3.33 831

seq\_name: /cgn2\_6/ptodata/1/iaa/5B\_COMB.ppt:US-08-613-220B-4

seq\_documentation\_block:  
: Sequence 4, Application US/08613220B  
: Patent No. 5958751  
: GENERAL INFORMATION:  
: APPLICANT: Murphy, Dennis  
: APPLICANT: Reid, John  
: TITLE OF INVENTION: ALPHA-GALACTOSIDASE  
: NUMBER OF SEQUENCES: 4  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Fish & Richardson, P.C.  
: STREET: 4225 Executive Square, Suite 1400  
: CITY: La Jolla  
: STATE: CA  
: COUNTRY: US  
: ZIP: 92037  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: Windows95  
: SOFTWARE: FastSeq for Windows Version 2.0  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/613,220B  
: FILING DATE: 08-MAR-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Haile, Ph.D., Lisa A.  
: REGISTRATION NUMBER: 38,347  
: REFERENCE/DOCKET NUMBER: 09010/004001  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 619-678-5070  
: TELEFAX: 619-68-5099  
: TELEX:  
: INFORMATION FOR SEQ ID NO: 4:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 346 amino acids  
: TYPE: amino acid  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: FRAGMENT TYPE: internal  
: US-08-613-220B-4  
  
alignment\_scores:  
Quality: 1580.00 Length: 364  
Ratio: 4.566 Gaps: 18  
Percent Similarity: 95.055 Percent Identity: 94.505  
  
alignment\_block:  
US-09-886-400-3 x US-08-613-220B-4 ..  
Align seg 1/1 to: US-08-613-220B-4 from: 1 to: 346  
1 TTGAGAGCGCTGCTCTTTTCACGGCAACCTCCAGTATGCGGAATCCCAA 50  
|||||  
1 LeuArgAlaLeuValPheHisGlyAsnLeuGlnTyrAlaGluIleProly 17  
51 GAGCGAAATCCCAAGGTCATAGAGAAGGCATACATCCAGTCATCGAGA 100  
|||||  
17 sSerGlu...ProLysValIleGluLysAlaTyrIleProValIleGluT 33  
101 CACTGATTAAAGAAGAAATCTCTTTTGGCTCAACATACGGGCTATACC 150  
|||||

```
33 hrLeuIleLysGluGlu...ProPheGlyLeuAsnIleThrGlyTyrThr 48
151 TTAAGTTCCTCCCGAAGGATATATTAGACCTCGTTAAAGGGGGCATCGC 200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 LeuLysPheLeuProLysAspIleIle...LeuValLysGlyIleAl 64
201 GAGTGACCTGATAGATATCGGACGAGCTACAGCGCAATACCTCC 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 aSerAspLeuIleGluIleGlyThrSerTyrThr...AlaIleLeuP 80
251 CCTCTCCCGCTTAGCAGAGTAGAGACACAAGTTCAGAGATAGGAA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
80 roLeuLeuProLeuSerArgValGluAlaGlnValGlnArgAspArg... 95
301 GTAAGGAAGAGCTCTCGAGCTTCTCAAGGGGATTCGTGCTCCAGA 350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96 ValLysGluGluLeuPheGluValSerProLysGlyPheTrpLeuProG 112
351 GCTGCGCTATGACCGGATATCCCTCCCTACTGAAGACACGGTTATG 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 uLeuAla...AspProIleIleProAlaIleLeuLysAspAsnGlyTyrG 128
401 AGTATCTATTTCGCGGAGGGGAGGCGATCTTCTCAGCTCATCTCAAC 450
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
128 luTyrLeuPheAlaAsp...GluAlaMetLeuPheSerAlaHisLeuAsn 143
451 TCGCGGATAAGGCAATTAACCGCTCTATCCACACCTTATAAGGCCCA 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
144 SerAlaIleLysProIleLysProLeu...ProHisLeuIleLysAlaG 159
501 AAGGAAACGGCTTTAGGTACATCAGCTATCTCCTTGCTCAGGGAGC 550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 nArgGluLysArgPheArgTyrIleSerTyrLeuLeu...LeuArgGluL 175
551 TTAGGAAGCGGATAAGCTCGTTTTTGAAGGTAAGCTAACGCTAAAGGCA 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 euArgLysAlaIleLysLeuValPheGluGlyLysValThrLeuLys... 190
601 GTCAAGACATCGAAGCGGACCGCTTTGGTGGCGGTGAACACGGCTGT 650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 ValLysAspIleGluAlaValProValTrpValAlaValAsnThrAlaVa 207
651 AATCTCGCGATCGGAGGCTCTCTTATGATCTCTAAGAAAGTGCGCA 700
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
207 lMetLeu...IleGlyArgLeuProLeuMetAsnProLysLysValAlaAs 223
701 GCTGGATAGGACACAGGACCAACATCTTCTATACGGCACCGCATATAG 750
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 erTrpIleGluAspLys...AsnIleLeuLeuTyrGlyThrAspIleGlu 238
751 TTCATTGCTATAGGACATTCGAGCTCAGAGATGATGATGTTGAGGGATT 800
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
239 PheIleGlyTyrArgAspIleAlaGly...ArgMetSerValGluGlyLe 254
801 ATTAGAGTTATAGACAGCTCAACTCGGAACGTGCTCCTTCCTCAGAGC 850
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 uLeuGluValIleAspGluLeuAsnSerGluLeuCys...ProSerGluL 270
851 TGAACACAGTGGAGGAGGCTCTACTTACGAGCTTCGAGTTGGGCACCA 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 euLysHisSerGlyArgGluLeuTyrLeuArgThrSerTrpAla... 285
901 GATAGAGCTTGAGCATATGAGAGGAGGACGAAGGGAACGACAGACTTAA 950
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 AspLysSerLeuArgIleTrpArgGluAspGluGlyAsnAlaArgLeuAs 302
951 TATGCTGTCTTACAAATATAGGGGGCGAAGCTCGCCCTTTAGCCGAGAAC 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 nMetLeu...TyrAsnMetArgGlyGluLeuAlaPheLeuAlaGluAsnS 318
1001 GCGATGCAAGGGGATGGAGCGCCCTCCTCAGAGGAGGCTGGATGCTTC 1050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
318 erAspAlaArgGlyTrp...ProLeuProGluArgArgLeuAspAlaPhe 333
```

```
1051 CGGGCGATATATACGATTCGAGGGTGAAATGGGAACCT 1092
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 ArgAlaIleTyrAsnAspTrpArgGly...AsnGlyGluPro 346
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-07-894-212A-8
seq_documentation_block:
; Sequence 8, Application US/07894212A
; Patent No. 5368893
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MUKAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/894, 212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-894-212A-8
alignment_scores:
Quality: 141.50 Length: 391
Ratio: 0.773 Gaps: 23
Percent Similarity: 46.803 Percent Identity: 22.762
alignment_block:
US-09-886-400-3 x US-07-894-212A-8
Align seg 1/1 to: US-07-894-212A-8 from: 1 to: 647
73 GAGAAGCATACATCCCGATCGACACTGATTAAGAAGAAATTC 122
27 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 41
123 T.....TTGGGCTCAACATACGGGCTATACCTTAAAGTTCTCCTCC 163
41 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 58
164 CG.....AAGGATATTATAGACTCGTTAAAGGGGCGCATCGGAGT 204
58 lAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 74
```

```
205 GACCTGATAGATAATCGGAACAGCTACAGCAGCAATCTCCCT 254
75 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAla 91
255 COTCCGCTTAGCAGAGTAGAAGACACAGTTCAGAGATAGGAAGTTA 304
91 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 105
305 AGGA.....GAGCTCTTCAGCTTCTCCAAAGGATCTGGCTGCCA 348
105 ysglutrPalalysSerIleGlyPheaspAlaArgGlyValrPLeuthr 121
349 GAGCTCGCTATGACCCGATAATCCCTGCCATACTGAAGACACGGTTA 398
122 GluArgValrPLeuProGluLeuValIysThrLeuLysGluSerGlyI 138
399 TGAGTATCTATTCGCCGACGGGAGGCGATCTTTCTCAGCTCATCTCA 448
138 eAspTyrValIleValAsp.....AspTyrHisPheM 149
449 ACTCGCGGATAAAGCAATTAACCGCTCTAT...CCACACCTTATAAG 495
149 etSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrTyrThrGlu 165
496 GCCCAAGGGA.....AAGCGCTTAGTA 521
166 AspGlyGlyGluValIleAlaValPheProIleAspGluLysLeuArg. 181
522 CATCAGCTATCTCCTGTCTCAGGAGCTTAGGAAGCGGATAAAGCTC. 570
182 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 196
571 .....GTTTTGAAGTAGTAACTAAGCGAGTC.....AAA 606
196 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHisAsp 212
607 GACATCGAAGCGGTACCGCTTTGGTGCCCTGAACACGGTGTAAATGCT 656
213 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 228
657 CGCATCGAAGCGCTCTCTATGATCTTAAGAAAGTGCAGCTGGA 706
229 .....GluLysGlyTrpL 233
707 TA.....GAGGACAAGCACAATCTT 729
233 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 249
730 CTATACGGCACCAGATAGAG.....TTCATTGGCTATAG 764
250 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyrLe 266
765 GGACATTGACGCTACAGAATGAGTGTGAGGATATTAGAGGTTATAG 814
266 uProIleAlaSerTyr.....PheGluMet.... 274
815 ACGAGCTCAACTCGGAGCTGCTCCCTCC..... 843
275 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 287
844 .....TCAGAGCTGAAGCAGCTGGA.....AGGA 869
288 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 304
870 GCTCTACTTACGACTTCGAGTTGGGACCCAGATAAGAGCTTGAGGATAT 919
304 gValPheValArgGlyIleTrp.....LysAsnPhe...PheT 317
920 GGAGAGGACGACGAGGAC.....GCAAGCTTAATATGCTGCTCC 960
317 yrlLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 333
```

```
961 TACAATATGAGGCGCAACTGCCCTTTAGCCGAGAACAGCGATCAAG 1010
334 LysLeuValArgAsn..... 339
1011 GGGATGGAGCCCTCCCTGAGAGAGGCTGGATCCCTCCGGCGGATAT 1060
340 .....ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 351
1061 ATACGAT.....TGAGAGGGT 1077
351 ysaAsnAspAlaTyrTiphisGly 358
seq_name: /cgr2_6/ptodata/1/laa/5A_COMB.pep:US-07-894-212A-2
seq_documentation_block:
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABRY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KORULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-894-212A-2
alignment_scores:
Quality: 141.50 Length: 391
Ratio: 0.773 Gaps: 23
Percent Similarity: 46.803 Percent Identity: 22.762
alignment_block:
US-09-886-400-3 x US-07-894-212A-2
Align seg 1/1 to: US-07-894-212A-2 from: 1 to: 649
73 GAGAAGGCATACATCCAGCTATCCAGACACTGATTAAGAGAAATTC 122
29 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 43
```

```
123 T.....TTTGGCTCAACATAAGCGGTATACCTTAAAGTTCCTCC 163
43 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 60
164 CG.....AAGGATATTATAGACTCGTTAAAGGGGSCATCGGAGT 204
60 lNAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 76
205 GACCTGATAGATAATCGAAGCAGGTACACGCACGCATACCTCCCTC 254
77 GlyGlnValGluIleValAlaAlaGlyPheTyrGluProValLeuAla 93
255 CTTCCCGCTTAGCAGAGTAGAAGCACAAAGTTCAGAGAGATAGGAA 304
93 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 107
305 AGGA.....GACCTCTTCGAGCTTTCTCCAAAGGATTCGGCTGCC 348
107 ySGLuTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeu 123
349 GAGCTCGCTATGACCCGGAATACCTCGCCATACCTGAAGACACG 398
124 GluArgValTrpGlnProGluLeuValLysThrLeuLysLysSerGly 140
399 TGAGTATCTATTCGCCAGCGGAGGCGATGCTTTCTCAGCTCATCT 448
140 eAspTyrValIleValAsp.....AspTyrHisPheM 151
449 ACTCGGCTAAGCCCAATTAACCGCTCTAT...CCACACCTTATAA 495
151 eSerAlaGlyLeuSerLysGluLysLeuTyrTrpProTyrTyrThr 167
496 GCCCAAGGGAA.....AAGCGCTTTAGGTA 521
168 AspGlyGlyGluValIleAlaValPheProIleAspGlyLysLeu 183
522 CATCAGCTATCTCTGCTCAGCGAGCTTAGGAGCGGATAAAGTC. 570
184 .....TyrLeuIleProPheArgProValAspLysValLeuGlu 198
571 .....GTTTGTGAGGTAGTACGGTAAAGGCAGTC.....AAA 606
198 euHisSerLeuIleAspGlyAspGluSerLysValAlaValPheHis 214
607 GACATCGAAGCGGTACCGTTTGCTGGTGGCGTGAACACGCTGTAT 656
215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpVal 230
657 CGGCATCGAAGGCTTCTCTATGAATCCTAAGAAAGTGGCGAGTGA 706
231 .....GluLysGlyTrpL 235
707 TA.....GAGGACAAAGGACAACATCTT 729
235 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeu 251
730 CPATACGCCACCGATATAG.....TTTATTGGCTATAG 764
252 LeuTyrThrGluTyrLeuGluLysTyrLysProArgGlyLeuValTyr 268
765 GCACATTCGAGCTACAAATGAGTGTGAGGATTATTAGAGTTATAG 814
268 uProIleAlaSerTyr.....PheGluMet.... 276
815 ACGAGCTCAACTCGGAAGTGTCCCTCC..... 843
277 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 289
844 .....TCAGCTGAAGCAGCTGA.....AGGGA 869
290 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyr 306
870 GCTCTACTTACGAGCTTCGAGTTGGCCACCAGATAAGAGCTTGAGG 919
```

seq\_name: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:US-07-893-928A-1

seq\_documentation\_block:  
; Sequence 1. Application US/07893928A  
; Patent No 5578479  
; GENERAL INFORMATION:  
; APPLICANT: LADERMAN, KENNETH  
; APPLICANT: ANFINSEN, CHRISTIAN  
; TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC  
; TITLE OF INVENTION: ARCHAEABACTERIUM  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN  
; STREET: 1100 NEW YORK AVENUE, N.W.  
; CITY: WASHINGTON, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Tape  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/893.928A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16773  
; REFERENCE/DOCKET NUMBER: 95470/C-1197  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 650 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-893-928A-1

alignment\_scores:  
Quality: 141.50 Length: 391  
Ratio: 0.773 Gaps: 23  
Percent Similarity: 46.803 Percent Identity: 22.762

alignment\_block:  
US-09-886-400-3 x US-07-893-928A-1 ..  
Align seg 1/1 to: US-07-893-928A-1 from: 1 to: 650



```
alignment_scores:
  Quality: 104.00
  Ratio: 1.316
  Percent Similarity: 51.299
  Length: 154
  Gaps: 7
  Percent Identity: 25.974
```

```
alignment_block:
US-09-886-400-3 x US-08-270-013B-2
Align seg 1/1 to: US-08-270-013B-2 from: 1 to: 329

28 CTCAGTATCCGAAATCCAAAGACGCAATCCCAAG..... 66
   ::::: ::::: ::::: ::::: :::::
130 MetThrThrValPheLysGluSerGlyPheProLysAsnArgVal11 146
   ::::: ::::: ::::: ::::: :::::
67 .....GTCATAGAGAAGCATACATCCAGTCATCGAGAC 103
   ::::: :::::
146 eGlyGlnSerGlyValLeuAspThrAlaArg.....PheArgThrP 160
   ::::: :::::
104 TGATTAAGAAGAAATCTTTGGGCTCAACATACGGGCTATACCTTA 153
   ::::: :::::
160 heValAlaGlu.....LeuAsnIleSer..... 168
154 AAGTCTCTCCGGAAGATATTAGACCTCGTTAAAGGGGCGATCGGAG 203
   ::::: :::::
169 .....ValLysAspValThrGlyPheValLeuGlyHisGlyAs 182
204 TGACCTGATAGAGATAATCGGAACGAGCTACACGCACCAATA.....C 247
182 pAspMetValProLeuValArgTy-SerTyAlaGlyIleProLeuG 199
248 TCCCTCCCTCCGCTTAGCAGTAGAAGACACAAAGTTCAGAGATAGG 297
199 LuLysLeuIleProLysAspArgLeuAspAlaIleValGluArgThr 215
298 GAAGTTAAGAAGAGCTCTTCAGCTTCTCCAAAGGGA.....TTCGTG 341
216 LysGlyGlyGlyGlyIleValAsnLeuLeuGlyAsnGlySerAlaTy 232
342 GCTGCCAGAGCTCGCTATGACCCGATATCCCTCCATCTGAGGACA 391
232 rAlaProAlaAlaSerLeuValGluMetValGluAlaIleLeuLys 249
392 AC.....GGTTAT 399
249 InArgIleLeuProAlaIleAlaTyLeuGluGlyIleGlyTy 265
400 GAGTATCTATTC 411
266 GluGlyIleTy 269

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-838-418-2

seq_documentation_block:
; Sequence 2, Application US/08838418
; Patent No. 5744342
; GENERAL INFORMATION:
; APPLICANT: Sogabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,418
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

seq\_name: /cgn2\_6/ptodata/1/iaa/65\_COMB.pep:US-09-228-986-72

seq\_documentation\_block:

; Sequence 72, Application US/09228986

; Patent No. 6359198

; GENERAL INFORMATION:

; APPLICANT: Strabala, Timothy

; APPLICANT: Nieuwenhuizen, Niels

; TITLE OF INVENTION: Compositions Isolated from Plant Cells

; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling

; FILE REFERENCE: 11000/1020

; CURRENT APPLICATION NUMBER: US/09/228,986

; CURRENT FILING DATE: 1999-01-12

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 72

; LENGTH: 910

; TYPE: PRT

; ORGANISM: Pinus radiata

US-09-228-986-72

alignment\_scores:

Quality: 93.00 Length: 388

Ratio: 0.479 Gaps: 25

Percent Similarity: 50.000 Percent Identity: 22.423

alignment\_block:

US-09-886-400-3/rev x US-09-228-986-72 ..

Align seg 1/1 to: US-09-228-986-72 from: 1 to: 910

1072 TCCATCGTTATATATCCCGGAGGATCCAGCTCCTCAGGGAGG 1023

200 SerLeuGlnTyrLeuSerLeuSerGln.....GlyAsnLeuGlnGlySe 214

1022 GCTCCCATCCCTTCGACGCTGTCGCTAAAAGGGAGGTCGCC 973

214 rValProSerGluLeuGlyArgLeuSerGlnLeu.....IleVal 227

972 CTCATATGTAGACACATTAAGTCTTCGCTCCCTCGCTCCTC 923

228 LeuAspLeuPheGlyAsnHisLeuThrGlyCysIle.ProSerSerLeu 244

922 TCCATATC.....CTCAAGCTCTTATCTGGTCCCACTCGAGTCGGT 879

244 erAsnCysThrAsnLeuGlnLeuLeuAsp.IleGlyAspAsnGlnLeu 260

878 AAGTAGAGCTCCCTCCACTGTGCTTCAGCT.....PheAsnCysThr 848

260 lGlyHisIleProSerHisLeuCysThrLysLysThrThrGlnLeuMet 277

847 ..CTGAGGAGGACAGTTCGAGTTCGAGTTCGCTCTATACCTCTAATA 800

277 yrLeuArgLeuGly..AlaAsnGlnLeuSerGlySerVal..... 289

799 ATCCCTACACTCATCTCTGAGCTGCAATGTCCTATACCAATGAAC 750

290 ...ProSerSerLeu.....PheAsnCysThrLysLeuGln 300

749 TCTATACGCTGCGGTATAGAGAATGTTGCTTCCTCTATCCAGCT 700

301 GluIleAlaLeuProTyrAsnGlnLeuSerGlyIleValProMetGluLe 317

699 CGGC.....ACTTCTTAGGA.....TTCA 680

317 uGlyLysLeuThrHisLeuGlnArgLeuPhePheGlyAsnTyrPheI 334

679 TTAGAGGA..AGCCTCCGATCGCGACATTACAGCGCTCTTCACGCC 633

334 leSerGlyAsnThrMetArgCysProIleLeuThrAlaLeuSerAsnCys 350

632 ACCAAACGGGTACGGCTTCGATGCTTTGACTGCCTTAGCGTTACCTT 583

331 SerAspLeuGlnTyrValAspLeuSerGluAsnAsnPheThrGlyGlnLe 367

582 ACCT.....TCAAAAACGAGCTTTATCGCCTCTCTAA 551

367 uProPheSerIleGlyHisLeuSerLys..LysLeuTyrHisLeu...As 382

550 GCTCCCTGAGACCAAGGAGATAGCTGATGCTACCTAAAGCGCTTTCCCTT 501

382 pleuGlySerAsnGlnLeuAlaGlyGluIleProAlaIleGlyAsnL 399

500 TGGGCTTTATAAGGTGTGATAGAGCGGTTTAATGGCTTTATCGCCGA 451

399 euser..... 400

450 GTTGAGATGAGCTGAGAAAACATCGCTCCCGTCGGCAATAGATACT 401

400 ..... 400

400 CATAACCGTTGCTTCAGTATGCGAGGATATATCGGGTCATAGCGGAGC 351

401 ..SerLeuThrPheLeuAsnLeuGlyArgAsnTyr....PheThrGlySerI 416

350 TCTGCGACGCAAGATCCCTT.....TGGAGAAAGCTCGAAGAGCTCTTC 307

416 le.....ProSerSerLeuIleMetLeuGlnLysLeuGluArgLeuTyr 430

306 CTT.....AACTTCCCTATC..... 292

431 MetAspSerAsnAsnLeuGlnGlyAsnIleProMetGluIleGlyGlnLe 447

291 ....TCTCTGAACCTTGTCTTCTACTCTGCTAAGCGGAGGAGGG.... 250

447 uLysSerLeuGlyLeuLeuTyrLeuSerGlyAsnAsnLeuSerGlyLysI 464

249 .....GAGTATTCGCTGCTGCTAGCTCGTTCCTGATTTATCTCTAT..... 211

464 leProAspPheValAlaAsnLeuGlnGlnLeuArgTyrLeuTyrLeuAsn 480

211 ..... 211

481 HisAsnGlnLeuSerGlyAspIleAsnAlaAsnLeuGlyLysCysValAs 497

210 ..CAGTCTACTCGGATGCCCTTTTACGAGGTCTATAATATCTCTGGG 162

497 nLeuLeuLeuLeuAspLeuSerTyrAsnLysLeuSerGlyHisIleProG 514

161 AGGAACCTTTAAGGTATAGCGCTTATGTTAGCCCAAGGAATTTCTTC 112

514 lngLLeuAlaGlyLeuAlaAsnLeuAla.....PheTyr 525

111 TTTAAT 106

526 PheAsn 527

seq\_name: /cgn2\_6/ptodata/1/iaa/66\_COMB.pep:US-09-061-709-2

seq\_documentation\_block:

; Sequence 2, Application US/090617093

; Patent No. 6297364

; GENERAL INFORMATION:

; APPLICANT: Chen, Yao-Tsung

; APPLICANT: Gure, Ali

; APPLICANT: Tsang, Solam

; APPLICANT: Stockert, Elisabeth

; APPLICANT: Jager, Elke

; APPLICANT: Knuth, Alexander

; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

; FILE REFERENCE: LUD 5538

; CURRENT APPLICATION NUMBER: US/09/061,709B

; CURRENT FILING DATE: 1998-04-17



```
563 .....SerAlaAr 565
242 CAATACT...CCOCTCCTCCGCTTAGCAGATAGAACACACAGTTCAG 288
|||||
565 9GlyThrTyArgAlaSerProAla..ArgGlyLeuGluGlyLeuLeu 581
289 AGAGATAGGAAGTAAGGAAGAGCTTCTCGAGCTTCTCCAAAGGGATT 338
|||||
582 GluThrArgProAlaSerProGlyLeuTyProHisGluProArgGly.. 597
339 CTGCTGCCAGAGCTCCCTATGACCGGATAATCCCTGCCATCTGAAG 388
|||||
598 ....LeuValAlaLeuGlyLeuTyProArgHisIleSerIleLeu... 611
389 ACAACGGTTATGATCTATTCGCCGACCGGGAG.....GGC 426
|||||
612 .....GluGlyLeuTyAlaArgGlyThrTyArgCysTySer 624
427 ATGCTTTCTCAGCTCATCTCACTCGCGGATAAAGCAATTAACCGCT 476
|||||
625 GlyLeuTyGlyLeuAsnLeuTySerThrHisArgPro...ArgGlyLe 640
477 CTATCCACACTTATA.....AAGGCCAAAGGGAAGCGCT 514
|||||
640 uTyAlaArgGlyIleLeuGluAlaArgGlySerGluArgSerGluArgS 657
515 TTAGTATCATCAGCTATCTCTGCTCTAGG...GAGCTTAGGAAGCG 561
|||||
657 erGluArgGlyLeuTyIleLeuGluLeuGluSerGluArgMetGluThr 673
562 ATAAAGCTCGTTTGAAGGTAGGTAAAGCTAAAGCGACTCAACACAT 611
|||||
674 ValAlaLeuProHisGluThrTyArgThrHisArgAlaSerProSerGl 690
612 CGAAGCGCTACCGCTTTGGTGGCCGCTCAACAGCGCTGTAATCTCGCA 661
|||||
690 uArgAlaLeuAlaIleLeuGluAlaLeu.....AlaLeuTySerGlyL 705
662 TCGGAAGCTTCCTCTTATGAATCTTAAGAAAGTGGCGAGCTGATAGAG 711
|||||
705 euGlyLeuTyProHisGluSerGluArgAlaLeuAla..... 718
712 GACAAGGACAACTCTCTATAGCGCACCGATATAGTTCATTGGCTA 761
|||||
719 .....SerAsnThr..... 721
762 TAGGACATTCGAGCTACAGATGAGTGTGAGGGATTATTAGAGTTA 811
|||||
722 .....TyArgSerGluArgValAlaLeuLeuGluGlyL 733
812 TAGAGAGCTCAACTCGGAAGTGTGCTTCCCTCAGAG.....CTGAAG 855
|||||
733 euAsnSerGluArgSerGluArgIleLeuGluSerGluArgGlyLeuAla 749
856 CACAGTGAAGGAGCTTACTTACGAGACTTCGAGTTGGGCACCATATA 905
|||||
750 SerProProHisGluLeuTy.....SerCysTySerMetGluTh 763
905 GAGCTTGAGGATATGAGAGAG..... 927
763 rGlyLeuAlaLeuAlaLeuGluGlyLeuTyMetGluThrGlyLeuSerG 780
928 .....GACGAAGGG 936
780 luArgGlyLeuTyGlyLeuLeuLeuGluHisIleSerSerGluArgAla 796
937 AACGCAAGACTTAATATCTCTCTCAATATGAGGGCGAAGTCCGCT 986
|||||
797 SerProGlyLeuAsnIleLeuGluThrHisArg.....AlaLe 809
987 TTTATCCGAGAACAGCGATGCAAGGGATGGAGCCCTCCCTCCCTGAGGA 1036
|||||
```

```
809 uAlaSerGluArgSer.....GluArg 817
1037 GCGTGGATGCTTCCGGCGCATATAACGATTGGAGGGTGAATGGG 1086
|||||
817 lyLeuAsnThrTyArgGlyLeuTyThrHis...ArgAlaSerAsnThr 832
1087 GAACCT 1092
833 ArgPro 834

seq_name: /cnp2_6/ptodata/1/iaa/5A_COMB.psp:US-08-418-893D-23

seq_documentation_block:
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-418-893D-23

alignment_scores:
Quality: 88.50 Length: 358
Ratio: 0.497 Gaps: 21
Percent Similarity: 49.721 Percent Identity: 19.553

alignment_block:
US-09-886-400-3 x US-08-418-893D-23 ..
Align seg 1/1 to: US-08-418-893D-23 from: 1 to: 2089

25 AACCTCCAGTATGCCGAATCCCAAGAGCGAATCCCAAGGTCATAGA 74
|||||
```

```

1057 SerPheGlnPheAlaAspValProAlaLysAsp.....ArgValThrAr 1071
75 GAAGCGTATACCCAGTCATCGACGACACTGATAAA.....GAAGAAA 118
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1071 gGlnGlyPhePheSerValIleAspAspAlaSerLysPheAlaGlnGlnL 1088
119 TTCCT.....TTTGGGCTCAACATAACGGGCTATACCTTA 153
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 euProGluIleLeuAsnSerPheGlySerLysIleAlaGlyAspAlaSer 1104
154 AAGTTCCTCCCGAAGGATATATAGACCTCGTTAAAGGGGCGATCCGAG 203
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1105 LysGluGlyProValAsnValLeuGlnVal.....GlyAlaLeuSerGl 1119
204 TGACCTG...ATACAGATATC.....GGAAGAGCTACACGCGCAA 244
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1119 YAspIleSerIleGluAspLeuGluLysAlaThrSerAlaAsnLysAspL 1136
245 TACTCCCTCCCTCCGCTTAGCAGATAGACACAACTTCAGAGAGAT 294
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1136 ysLeuAsnMetLeuGlyValArgThrValThrAlaLeuIleProArgGly 1152
295 AGGAAGTTAAGAAAGCTCTTCGAGCTTCTCCAAAGGATTCGTGCT 344
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 LysLys.....AspProSerTyrTyrSerPh 1161
345 GCCAGAGCTCGCC.....TATGACCCGATAATCCCTGCCCATACTGA 385
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1161 eProGlnCysSerGlyPheLysGluAspProLeu..... 1172
386 AGNACACGGTTATGATATCTATTCCGCGAGGGGAGCGGATGCTTTC 435
1172 ..... 1172
436 TCAGCTCATCTCAACTCGGCGATAAGCAATTAACCGCTATCCACCA 485
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1173 .....ArgArgGlyMetArgProThrPheHisH 1182
486 CTTTATAAGGCCCAAGG...GAAAGGCGTTT..... 516
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1182 sLeuLeuGluLeuGlyArgLeuGluGluAsnPheAlaLeuGluArgIleP 1199
517 .....AGTACATCAGCTATCTCTGTCTCAGGAGCTT... 552
1199 roAlaValGlyArgAsnValGlnIleTyrValGlySerGluLysThrAla 1215
553 ...AGGAAGCGATAAGCTCGTCTTTTGAAGTAAGTAACGCTAAAGCG 599
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1216 ArgArgAsnAlaAlaGlnValPhe.....LeuArgAl 1227
600 AGTCAAGACATCGAAGCCGATCCCGTTGGTGGCGCTGAACACGGCTG 649
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1227 aIleSerHisThrProGlyLeuThrThrPheSerGlyAlaArgAlaAl 1244
650 TAATGCTCGCATCGGAAGCTTCTCTTATGATCTTAAGAAAGTGGCG 699
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1244 euLeuGlnGlyLeuAspGluLeuGluArgAlaGlnAlaAsnSerLysVal 1260
700 AGCTGGATAGGACAGACACATCTCTATACGGC..... 738
|||:|||||:|||||:|||||:|||||:|||||:|||||:
1261 Ser...ValGlnSerSerArgIleTyrLeuHisSerLeuProGluL 1276
739 .ACCGATATA.....GAGTTTCATTGGCTATAGG 766
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1276 nSerAspAlaThrProGluGluIleAlaLysGluPheGluGly..... 1290
767 ACATTGAGGCTACAGATAGTGTTCAGGGATTATTAGAGTTATAGAC 816
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1291 .....ValIleAsp 1293
817 GAGCTCAACTCGGAACGTGCTTCCCTCAGAGCTGAAGCAGACGTGAAG 866
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1294 LysLeuLysSerArgLeu.....AlaGlnAr 1302

```

```

867 GGAGCTCTACTTACGAGCTTCGAGTTGGCCACCATAGAGACTTGAGGA 916
| |||||:|||||:|||||:|||||:|||||:|||||:
1302 gLeuThrLysLeuArgValAspGlu.....IleGluThrLysValArgV 1317
917 TATGGAGAGAGGACGAA...GGGACGCGCAAG.....CTTAATATG 954
:|||||:|||||:|||||:|||||:|||||:|||||:
1317 alThrValGlnAspGluAspGlySerProArgValValProValArgLeu 1333
955 CTGCTCTACAATATAGGGGCGCAA 978
:|||||:|||||:|||||:|||||:|||||:|||||:
1334 ValAlaSerSerMetGlnGlyGlu 1341
seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-418-893D-24
seq_documentation_block:
; Sequence 24, Application: US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; US-08-418-893D-24

```

```

alignment_scores:
  Quality: 88.50      Length: 358
  Ratio: 0.497      caps: 21
  Percent Similarity: 49.721      Percent Identity: 19.553

alignment_block:
  US-09-886-400-3 x US-08-418-893D-24

```

Align seg 1/1 to: US-08-418-893D-24 from: 1 to: 2089

```
25 AACCTCCAGTATCGCGAAATCCCAAGAGAGCAATCCCAAGGTCTATAGA 74
   ::::::::::::::::::::
1057 SerPheGlnPheAlaAspValProAlaLysAsp.....ArgValThrAr 1071
   ::::::::::::::::::::
75 CAAGGCATATACCCAGTCTCATCCAGACACTGATTA.....GAAGAA 118
   ::::::::::::::::::::
1071 gGlnGlyPhePheSerValIleAspAspAlaSerLysPheAlaGlnGlnL 1088
   ::::::::::::::::::::
119 TTCTCT.....TTTGGGTCAACATAACGGGCTCTATACCTTA 153
   ::::::::::::::::::::
1088 euProGluIleLeuAsnSerPheGlySerLysIleAlaGlyAspAlaSer 1104
   ::::::::::::::::::::
154 AAGTTCTCCCGAAGGATATATAGACCTCGTTAAAGGGGGCATCCGAG 203
   ::::::::::::::::::::
1105 LysGluGlyProValAsnValLeuGlnVal.....GlyAlaLeuSerGly 1119
   ::::::::::::::::::::
204 TGACCTG...ATAGAGATAATC.....GGAAGAGACTACACGACGCA 244
   ::::::::::::::::::::
1119 yAspIleSerIleGluAspLeuGluLysAlaThrSerAlaAsnLysAspL 1136
   ::::::::::::::::::::
245 TACTCCCTCTCTCCCTAGCAGTATAGACACAAAGTTTCAGAGAT 294
   ::::::::::::::::::::
1136 ysLeuAsnMetLeuGlyValArgThrValThrAlaLeuIleProArgGly 1152
   ::::::::::::::::::::
295 AGGAAGTTAAGGAAGAGCTCTCCGCTTCTCCAAAGGCAATCTGGCT 344
   ::::::::::::::::::::
1153 LysLys.....AspSerSerTyrTyrSerPh 1161
   ::::::::::::::::::::
345 GCCAGAGCTGCC.....TATGCCGATTAATCCCTGCCATCTGA 385
   ::::::::::::::::::::
1161 eProGlnCysSerGlyPheLysGluAspProLeu..... 1172
   ::::::::::::::::::::
386 AGGACACGGTTATGAGTATCTATTCGCCGACGGGGAGGCGATCTTTC 435
   ::::::::::::::::::::
1172 ..... 1172
436 TCAGCTCATCTCAACTCGCGGATAAAGCCAAATTAACCGCTCATCCACA 485
   ::::::::::::::::::::
1173 .....ArgArgGlyMetArgProThrPheHisH 1182
   ::::::::::::::::::::
486 CCTTATAAGCCCAAGG...GAAAAGCGCTT..... 516
   ::::::::::::::::::::
1182 sLeuLeuGluLeuGlyArgLeuGluGluAsnPheAlaLeuGluArgIleP 1199
   ::::::::::::::::::::
517 .....AGGTACATCAGCTATCTCTGCTCAGGAGCTT... 552
   ::::::::::::::::::::
1199 toAlaValcIlyArgAsnValGlnIleTyrValGlySerGluLysThrAla 1215
   ::::::::::::::::::::
553 ...AGGAAGCGGATAAGCTCGTTTTCAGAGGTAAAGTAAAGCTAAGGC 599
   ::::::::::::::::::::
1216 ArgArgAsnAlaAlaGlnValPhe.....LeuArgAl 1227
   ::::::::::::::::::::
600 AGTCAAGACATCGAAGCGTACCGCTTTGGTGGCGGTGAACACAGCGTG 649
   ::::::::::::::::::::
1227 aIleSerHisThrProGlyLeuThrThrPheSerGlyAlaArgAlaL 1244
   ::::::::::::::::::::
650 TAATGCTCGGATCGGAAGGCTCTCTTATGATCTCTAAGAAGTGGCG 699
   ::::::::::::::::::::
1244 euLeuGlnGlyLeuAspGluLeuGluArgAlaGlnAlaAsnSerLysVal 1260
   ::::::::::::::::::::
700 AGCTGGATAGAGACAGCAACATCTTCTATACGCG..... 738
   ::::::::::::::::::::
1261 Ser...ValGlnSerSerArgIleTyrLeuHisSerLeuProGluL 1276
   ::::::::::::::::::::
739 ACCGATATA.....GAGTTCATTGGCTATAGG 766
   ::::::::::::::::::::
1276 nSerAspAlaThrProGluGluIleAlaLysGluPheGluGly..... 1290
   ::::::::::::::::::::
767 ACATTCCAGGCTACAGATGAGTCTTGAGGGATTATTAGAGTTATAGAC 816
   ::::::::::::::::::::
1291 .....ValIleAsp 1293
```

```
817 GAGCTCAACTCGGAAGTGTGCTTCCTCCCTCAGAGCTGAACACAGTGGAG 866
   ::::::::::::::::::::
1294 LysLeuLysSerArgLeu.....AlaGlnAr 1302
   ::::::::::::::::::::
867 GGAGCTCTACTTACGGACTTCGAGTTGGGCACGACAGATAAGAGCTTGAG 916
   ::::::::::::::::::::
1302 gLeuThrLysLeuArgValAspGlu.....IleGluThrLysValArgV 1317
   ::::::::::::::::::::
917 TATGGAGAGAGGACGAA...GGGACGCAAGA.....CTTAATATG 954
   ::::::::::::::::::::
1317 alThrValGlnAspGluAspGlySerProArgValValProValArgLeu 1333
   ::::::::::::::::::::
955 CTGTCCTACATATATAGGGCGCAA 978
   ::::::::::::::::::::
1334 ValAlaSerSerMetGlnGlyGlu 1341
   ::::::::::::::::::::
```

seq\_name: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:US-09-355-166-1

```
seq_documentation_block:
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
; US-09-355-166-1
```

```
alignment_scores:
Quality: 87.50 Length: 333
Ratio: 0.557 Gaps: 19
Percent Similarity: 47.147 Percent Identity: 22.823

alignment_block:
US-09-886-400-3 x US-09-355-166-1 ..
Align seg 1/1 to: US-09-355-166-1 from: 1 to: 657
```

```
84 CATCCAGTCTATCGAGACTGTATTAAGAGAAATTCCTTTTGGGCTCA 133
   ::::::::::::::::::::
61 HisGlyGluLysArgSerThrAspProArgTyrSer.Pro..... 73
   ::::::::::::::::::::
134 ACATACCGGCTATACCTTAAGTTCCTCCCGAG.....GATATT 174
   ::::::::::::::::::::
74 .....AspGlyArgThrLeuAlaPheIleSerAspArgGluGlyAspAla 88
   ::::::::::::::::::::
175 ATAGACCTC.....GTTAAGGGGCGATCCGAGTACCTGAT 212
   ::::::::::::::::::::
89 AlaGlnLeuTyrIleMetSerThrGluGlyGluAlaArgLysLeuTh 105
   ::::::::::::::::::::
213 AGAGATA...ATCGAGACGAGCTACAG.....CACG 241
   ::::::::::::::::::::
105 RasPileProTyrGlyValSerLysProLeuTyrSerProAspGlyGlu 122
   ::::::::::::::::::::
242 CAATATCTCCCTCTCTCCGCTTAGCAGAGTAGACCAAGTTCAGAGA 291
   ::::::::::::::::::::
122 erIleLeuValThrIleSerLeuGlyGluSerIleAspAspArg 138
   ::::::::::::::::::::
292 GATAGGGAAGTTAGGAA.....GAGCTCTTCGAGCTTC 326
   ::::::::::::::::::::
139 GluLysThrGluGlnAspSerTyrGluProValGluValGlnGlyLeu 155
   ::::::::::::::::::::
327 TCCAAG.....GGATTCTGGCTGCCAGAGCTCGCTATGACCCGATAA 370
   ::::::::::::::::::::
```

```

155 rTyrLysArgAspGlyLysGlyLeuThrArgGlyAlaTyrAlaGlnLeuV 172
371 TCCTGCCATCTAGAG..... 387
172 alLeuValSerValLysSerGlyLysGlyLysGlyLeuThrSerHisLys 188
388 ...GACACGGTTATGATGATCTATTCGCGGAGGAGCGATCTTT 434
189 AlaAspHisGlyAspProAlaPheSerProAspGlyLysTrpLeuValph 205
435 CTGAGCTCATCTC.....AACTCGGCGATAAGCCA..... 465
205 eSerAlaAsnLeuThrGluThrAspAlaSerLysProHisAspValt 222
466 .....ATTAAACGGCTCTATCCACAC... 486
222 yrileMetSerLeuLysGlyLysPheLysGlnValThrProHisArg 238
487 .....CTTAAAGCCCAAGGAAAGCGCTTTAGGTACATCAGCTA 530
239 GlySerPheGlySerSerPheSerProAspGlyArgTyrLeuAlaLe 255
531 TCCTCTGGTCTCAGGAGCTTAGGAGGCG.....A 562
255 uLeuGlyAsnGluLysGluTyrLysAsnAlaThrLeuSerLysAlaTrpL 272
563 TAAAGCTCGCTTTTGAAGGTAAAGTAACG...CTAAAGCGAGCTCAAGAC 609
272 euTyrAspIleGluGlnGlyArgLeuThrCysLeuThrGluMetLeuAsp 288
610 ATCGAACCCGTACCCCTTGGGTGGCGGTGAACACGGCTGTATGCTCGG 559
289 ValHisLeuAlaAsp.....AlaLeuIleGlyAspSerLeuIleGlyL 303
660 CATCGGAAGCTTCCTCTTATGATTAATCCTAAGAAAGTGGCGAGCTGGATAG 709
303 yAlaGluGlnArgProIle.....TrpThrL 312
710 AGGACAGGACACATCTCTCTATACGGCACCGCATATAGATTCATGGC 759
312 ysAspSerGlnGlyPheTyrValIleGlyThrAspGlnGlySerThrGly 328
760 TATAGGACATTCAGGCTACAGAATGAGTGTGGAGGATATTAGAGGT 809
329 Ile.....TyrTyrIleSerIleGluGlyLeuValTyrPr 340
810 TATAGACAGCTCAACTCGAAGCTGCGCTTCCTCAGAGCTGAGGCACA 859
340 oIle...ArgLeuGlnLysGlu..... 346
860 GTGGAAGGGAGCTCTACTTACGAGCTTCGAGTTGGGCACCATAG 906
347 .....TyrIleAsnSerPheSerLeuSerProAspGlu 357

```

seq\_name: /cqn2\_6/ptodata/1/1aa/5B\_COMB.pep:US-09-141-135-2

seq\_documentation\_block:

; Sequence 2, Application US/09141135

; Patent No. 5961729

; GENERAL INFORMATION:

; APPLICANT: CHUN, Jong Yoon

; APPLICANT: LEE, Yong Yoon

; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisic

; FILE OF INVENTION: Acid Isolated from Arabidopsis thaliana

; FILE REFERENCE: 1942/31

; CURRENT APPLICATION NUMBER: US/09/141,135

; CURRENT FILING DATE: 1998-08-27

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: WordPerfect 6.1/Windows

; SEQ ID NO 2

; LENGTH: 235

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-141-135-2

alignment\_scores:

Quality: 87.00 Length: 233

Ratio: 0.763 Caps: 11

Percent Similarity: 48.927 Percent Identity: 21.030

alignment\_block:

US-09-886-400-3 x US-09-141-135-2 ..

Align seg 1/1 to: US-09-141-135-2 from: 1 to: 235

```

493 AAGGCCCAAGGAAAGCGCTTTAGGTACATCAGCTATCTCTGGTCT 542
|||||..... 35
27 LysSerAsnAsnGlnLysArgPheAsn..... 35
543 CAGGAGCTTAGGAGCGCATAAAGCTCTCTTTTGAAGTAAGTAAGCGC 592
|||||..... 1
36 .GluGluGlnIleLysSerLeuGluLeuIlePheGluSerGluThrArgL 52
593 TAAAGCAGTCAAGACATCGAAGCGGTACCCGTTTGGGTGGCGGTGAAC 642
|||||..... 59
52 euLysProArgLysLysValGln..... 59
643 ACGGCTGTATGCTCGGATCGGAGGCTTCCTCTATGATCTCTAAGAA 692
|||||..... 70
60 ValAlaArgGluLeuGly.....LeuGlnProArgL 70
693 AGTGGGAGCTGGATAGAGGACAAAGGAC..... 720
70 nMetThrIleTrpPheGlnAsnLysArgAlaArgTrpLysThrLysGlnL 87
721 .....AACATTCTCTTATACGGCACCGCATATAGATTCATT 756
|||||.....Ala 96
87 euGluLysGluTyrAsnThrLeuArg..... 96
757 GGTATAGGACATTCAGGC.....TACAGAATGAGTGTGA 794
97 AsnTyrAsnAsnLeuAlaSerGlnPheGluIleMetLysLysGluLysG 113
795 GGGATATTAGAGTTATAGACGAGCTCAACTCGGAACGTGCTTCCTCC 844
113 nSerLeuValSerGluLeuGlnArgLeuAsnGluGluMetGlnArgProL 130
845 CAGAGCTGAGCAC.....AGTGAAGGAGGCTCTACTACGG 882
130 yGluGluLysHisHisGluCysGlyAspGlnGlyLeuAlaLeuSer 146
883 ACTTCGAGTTGGGCACCATGAT...AAGAGCTTTCAGGATATGAGAGAG 929
147 SerSerThrGluSerHisAsnGlyLysSerGluProGluGlyArgLeuAs 163
930 CGAAGGACCGCAAGACTTAATATGCTGCTCCACAT..... 966
163 pGlnGlySerValLeuCysAsnAspGlyAspTyrAsnAsnAsnIleLysT 180
967 .....ATCAGGGGC.....CAACTCGCCCTTTTA 990
180 hrGluTyrPheArgValGlnGlyGluThrAspHisGluLeuMetAsnIle 196
991 GCGGAGAACCGCATATATACGATTGGAGGGGTGAAATGGGAA 1089
197 ValGluLysAlaAsp..... 201
1041 GGATCGCTTCGCGGCGATATATACGATTGGAGGGGTGAAATGGGAA 1089
202 .AspSerCysLeuThrSerSerGluAsnTrpGlyGlyPheAsnSerAsp 217

```

seq\_name: /cqn2\_6/ptodata/1/1aa/5A\_COMB.pep:US-08-248-021A-2

seq\_documentation\_block:

; Sequence 2, Application US/08248021A



```

; Patent No. 5648240
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Jonsson, Klas
; APPLICANT: Patti, Joseph M.
; APPLICANT: Gurusiddappa, Sivashankarappa
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/248,021A
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: TAMK:155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 689 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-248-021A-2

alignment_scores:
    Quality: 86.50      Length: 271
    Ratio: 0.765        Gaps: 11
    Percent Similarity: 41.697    Percent Identity: 18.819

alignment_block:
US-09-886-400-3 x US-08-248-021A-2

Align seg 1/1 to: US-08-248-021A-2 from: 1 to: 689

97 GAGACACTGATTAAAGAGAATTCCTTTGGCTCAACATACGGGCTA 146
||||| :||||| :||||| :||||| :||||| :||||| :|||||
264 GluAlaLysValAsnGlnValProTyrSerIleAsnLeuAsnGlyTh 280
147 TACCTTAAAGTTCCTCCCG..... 165
||||| :||||| :||||| :||||| :||||| :||||| :|||||
280 rThrThrAsnIleGlnSerAsnLeuAlaPheSerAsnLysProThrP 297
166 .....AAGGATATTAGACCTCGTTAAAGGGCGCATCGCGAGTAC... 207
||||| :||||| :||||| :||||| :||||| :||||| :|||||
297 snTyrLysAsnLeuThrThrLysValLysSerValLeuLysSerArg 313
207 ..... 207
314 GlyValSerGluArgAspLeuLysHisAlaLysAlaTyrTyrThrVa 330
208 .....CTGATAGATAATCGGAACGAGCT 232
330 lTyrPheLysAsnGlyGlyLysArgValIleHisLeuAsnSerAsnIle 347
233 ACAGCGACCAATATCCCTCCCTCCGCTTACGAGTAGAAGACAA 282
||||| :||||| :||||| :||||| :||||| :||||| :|||||
347 yrThrAlaAsnLeuValHisAlaLysAspValLysArgIleGluValThr 363

```

```

283 GTTCAGACAGATAGGAAGCTTAAGGAAGAGCTC.....TTCGAGCT 323
||||| :||||| :||||| :||||| :||||| :||||| :|||||
364 ValLysThrValSerLysValLysValLysAlaGluArgTyrValProTyrThr 380
324 TTCCTCCAAAGGATTCTGGCTGCGCAGAGCTC..... 354
||||| :||||| :||||| :||||| :||||| :||||| :|||||
380 eAlaValAsnGlyAlaSerAsnProThrLeuSerAspLeuLysPheThrG 397
355 .....GCCTATGACCCGATA.....ATCCTGTGCCATA 391
||||| :||||| :||||| :||||| :||||| :||||| :|||||
397 lYAspSerArgValSerTyrSerAspIleLysLysLysValLysSerVal 413
382 CTGAAG...GACAAACGGT..... 396
||||| :||||| :||||| :||||| :||||| :||||| :|||||
414 LeuLysHisAspArgGlyIleGlyGluArgGluLeuLysTyrAlaGluL 430
397 .....TATGAGTATCTATTCGCCGACGGGAGGCGATCCTTTTCTCAG 439
||||| :||||| :||||| :||||| :||||| :||||| :|||||
430 sAlaThrTyrThrValHisPheLysAsnGlyThrLysLysValIle... 445
440 CTCATCTCAACTCGCGATAAGCCAATTAAACCGCTCTATCCACACCTT 489
||||| :||||| :||||| :||||| :||||| :||||| :|||||
446 ..AsnLeuAsnSerAsnIleSerGlnLeuAsnLeuLysValLysAsp 461
490 ATAAAG.....GCCCAAAGGGAAGAGCGCTT 515
||||| :||||| :||||| :||||| :||||| :||||| :|||||
462 lLeLysAsnIleAspIleAspValLysThrGlyAlaLysAlaLysVal 478
516 TAGGTACATCAGCTAT..... 531
478 rSerTyrValProTyrThrIleAlaValAsnGlyThrThrProIleA 495
532 .....CTCCTTGGTCTCAGGGAGCTT 552
||||| :||||| :||||| :||||| :||||| :||||| :|||||
495 lAsrLysLysLeuLysSerAsnLysGlnLeuIleGlyTyrGlnAspLeu 511
553 AGGAAGCGCATAAAGCTCGTTTGAAGGTAAGGTAAAGCTAAAGGCGAGT 602
||||| :||||| :||||| :||||| :||||| :||||| :|||||
512 AsnLysLysValLysSerValLysHisAsp.....ArgGlyI 525
603 CAAGACATCGAA 615
525 eAsnAspIleGlu 529

```

seq\_name: /cgn2\_5/ptodata/1/laa/6B\_COMB.pep:US-08-910-925-3

```

seq_documentation_block:
; Sequence 3, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:

```



---

OM of: US-09-886-400-3 to: A\_Geneseq\_032802.\* out\_format : pfs  
 Date: Jun 11, 2002 10:33 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-WDEL-frame1\_n2p.model -DRV=xlh  
 -Q/cgn2\_1/USPTO\_spool/US09886400/runat\_11062002\_150913\_17285/app\_query.fasta.1.1169  
 -DB-A\_Geneseq\_032802 -QFM=fastan -SUFFIX=rag -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
 -TRANSHUMAN40.cdi -LIST=7.45 -DOCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs  
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09886400.@CGL1\_1\_0 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-886-400-3  
 Query length: 1095  
 Database: A\_Geneseq\_032802.\*  
 Database sequences: 747574  
 Database length: 11107396  
 Search time (sec): 53.92000

## score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW34643 +			1873.00	3855.49	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			145.50	264.50	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			143.50	260.46	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			124.00	211.28	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			121.50	214.74	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			104.50	186.48	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			104.50	185.97	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			104.00	180.01	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			104.00	180.01	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			104.00	177.55	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			104.00	177.19	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			103.50	173.16	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			103.50	172.18	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			97.00	159.76	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			96.50	159.78	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			96.00	155.99	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			96.00	155.21	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			94.50	163.94	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			93.50	165.15	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			93.00	161.71	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			93.00	144.58	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			93.00	140.31	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			92.50	138.20	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.50	148.00	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.00	166.55	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.00	165.45	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.00	165.56	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.00	155.25	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			92.00	144.42	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			91.50	159.38	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			91.50	156.29	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAW34643 +			91.00	160.58	
/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW34643 +			91.00	159.82	

/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW80176 + 91.00 159.82 0.2940  
 /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW82559 + 91.00 159.82 0.2940  
 /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW82373 + 91.00 157.99 0.3182  
 /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW62551 - 91.00 152.72 0.3998  
 /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AAW51656 + 91.00 148.33 0.4833

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:AAW34643

## seq\_documentation\_block:

ID AAW34643 standard; Protein; 364 AA.  
 AC AAW34643;  
 DT 27-MAR-1998 (first entry)  
 DE Thermostable alpha-galactosidase AEDII12RA-alpha-gal-18GC.  
 KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;  
 KW food processing; alpha glycoside hydrolysis; raffinose;  
 KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-18GC.  
 OS Thermococcus alcaliphilus strain AEDII12RA.

Key Location/Qualifiers  
 Misc-difference 329 /note= "encoded by CTT"

WO9732974-A1.

12-SEP-1997.

05-FEB-1997; 97WO-US01452.

08-MAR-1996; 96US-0613220.

(RECO-) RECOMBINANT BIOCATALYSIS INC.

Murphy D, Reid J, Rudolph MJ;

WPI: 1997-470541/43.

N-PSDB: AAT93753.

Nucleic acid encoding alpha-galactosidase from Thermococcus

alcaliphilus - used in food processing to hydrolyse

alpha-glycosides, e.g. raffinose

Claim 1; Fig 1; 32pp; English.

This protein comprises AEDII12RA-alpha-gal-18GC, a claimed thermostable alpha-galactosidase of Thermococcus alcaliphilus AEDII12RA, a bacterium that shows optimum growth at 85 deg C and pH 9.5. Also claimed are: (1) an isolated polynucleotide (see AAT93753) encoding the alpha-galactosidase; (2) a vector containing the polynucleotide or homologous or complementary sequences; (2) host cells containing the vector; (3) a process for producing the alpha-galactosidase in transformed or transfected host cells; an enzyme showing at least 70% identity to alpha-galactosidase and comprising at least 30 amino acid residues of its sequence; and (4) a method for hydrolysing alpha-galactoside bonds using the enzyme. The enzyme can be used to hydrolyse raffinose to sucrose and glucose in sugar beet processing (raffinose inhibits crystallisation of sucrose), and as a digestive aid to hydrolyse raffinose, stachyose and verbascose in beans and other gassy foods.

Sequence 364 AA;

alignment\_scores:  
 Quality: 1873.00 Length: 364  
 Ratio: 5.146 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.725  
 alignment\_block:

US-09-886-400-3 x AAW34643 ..

Align seg 1/1 to: AAW34643 from: 1 to: 364

```
1 TTGAGAGCGCTCGCTTTCACGGCAACCTCCATGTCGCGAAATCCCAA 50
1 LeuArgAlaLeuValPheHisGlyAsnLeuGlnTyAlaGluLeuProly 17
51 GAGCGAATCCAAAGGTCATAGAGAGGAGCATACCCAGTCATCGAGA 100
17 sSerGluileProlysValileGluLysAlaTyriileProValileGluT 34
101 CACTGATTAAAGAAATTCCTTTGGCTCAACATAACGGGCTATACC 150
34 hrLeuileLysGluGluileProPheleLysLeuasnlethrGlyTyThr 50
151 TTAAGTTCCTCCGAGGATATATAGACTCGTTAAAGGGGCGCATCCG 200
51 LeuLysPheLeuProLysAspIleleAspLeuValLysGlyIleAl 67
201 GAGTGACCTGATAGAGATAATCGGAAGAGCTACACGCGCAATACTCC 250
67 aserAspLeuileGluileleGlyThrSerTyThrHisAlaileLeuP 84
251 CCCTCCTCCCGCTTAGCAGAGTAGAAGCACAACTTCAGAGAGATAGGAA 300
84 roLeuLeuProLeuSerArgValGluAlaGlnValGlnArgAspArgGlu 100
301 GTTAAAGAGAGCTCTCGAGCTTCTCCAAAGGATTCGGGTCGCCAGA 350
101 ValLysGluLeuLeuPheGluLeuSerProLysGlyPheTrpLeuProgl 117
351 GCTCGCTATGACCCGATAATCCTCGCATCTGAAGGACAAAGCGTTATG 400
117 uLeuAlaTyAspProileleProAlaileLeuLysAspAsnGlyTyG 134
401 AGTATCTATTCGCGCGGAGGCGGAGTGTCTTCAGCTCATCTCAAC 450
134 LuTyLeuPheAlaAspGlyGluAlaMetLeuPheSerAlaHisLeuAsn 150
451 TCGGCGATAAGCAATTAACCGCTCTATCCACACCTTATAAAGGCCCA 500
151 SerAlaileLysProileLysProLeuTyProHisLeuileLysAlaGl 167
501 AAGGAAAGCGTTTAGTACATACATCTCCTTGGTCTCAGGAGC 550
167 naRgGluLysargPheArgTyriileSerTyLeuLeuGlyLeuArgGluL 184
551 TTAGAAGGCGATAAAGCTCGTTTTGAAGGTAAAGTAAAGGCA 600
184 euArgLysAlaileLysLeuValPheGluLysValThrLeuLysAla 200
601 CTCAAAGACATCGAAGCGCTACCGCTTGGGTGGCGCTGAACACGCTGT 650
201 ValLysaspileGluAlaValProValTrpValAlaValAsnThrAlaVa 217
651 AATCGTCGGCATCGAAGGCTTCTCTTATGAATCTTAAGAAAGTGGCA 700
217 lMeLeuGlyileGlyArgLeuProLeuMetAsnProLysLysValAlas 234
701 GCTGGATAGGACAAAGCAACATCTTCTATACGGCACCGCATATAGAG 750
234 erTpIleGluAspLysAspAsnileLeuLeuTyrglyThrAspIleGlu 250
751 TTCATTGGCTATAGGACATGAGGCTACAGATGAGTGTAGGGAT 800
251 PheileGlyTyArgAspIleAlaGlyTyArgMetSerValGluGlyLe 267
801 ATTAGAGTTATAGAGAGTCAACTCGGAAGTGTGCTTCCCTCCAGAGC 850
267 uLeuGluValileAspGluLeuAsnSerGluLeuCysLeuProSerGluL 284
851 TGAAGCACAGTGGAGGAGCTCTACTACGGACTTCGGAGTGGGCACCA 900
```

```
|||||
284 euLysHisSerGlyArgGluLeuTyriLeuArgThrSerSerTrpAlaPro 300
901 GATAAGAGCTTGAGGATATGAGAGAGGAGCAAGAGGCAAGACTTAA 950
|||||
301 AspLysSerLeuArgIleTrpArgGluAspGluGlyAsnAlaArgLeuAs 317
951 TATCTCTCTCTACATATAGAGGGCGAACTCGCCCTTTTAGCGGAGACA 1000
|||||
317 nMeLeuSerTyAsnMetArgGlyGluLeuAlaPheLeuAlaGluAsnS 334
1001 GCGATGCAAGGGATGGAGCCCTCCCTCCAGAGAGGCTGGATGCTTC 1050
|||||
334 erAspAlaArgGlyTrpGluProLeuProGluArgArgLeuAspAlaPhe 350
1051 CGGCGGATATATACGATTGGAGGGGTGAATGGGGAACCT 1092
|||||
351 ArgAlaileTyAsnAspTrpArgGlyGluAsnGlyGluPro 364
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA96088
seq_documentation_block:
ID AAB96088 standard; Protein: 655 AA.
XX
AC AAB96088;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi alpha-amyase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX
PN FR2792651-Al.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99FR-0005034.
XX
PR 21-APR-1999; 99FR-0005034.
XX
(CNRS ) CNRS CENT NAT RECH SCI
(PFRE-) IFREMER INST FR RECH EXPL MER.
XX
Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querallou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
proteins useful in industry -
XX
Claim 7; Pages 711-712; 1657pp; French.
XX
The present invention relates to the genomic sequence of Pyrococcus
abyssi (see AAF86431 and AAF41223-7) and P. abyssi proteins. P. abyssi is
a hyperthermophilic archaeon, which is isolated from deep-sea P. abyssi
hydrothermal vents. The present sequence is one such P. abyssi protein.
The proteins of the present invention have various potential industrial
uses, since the proteins are stable at very high temperatures, some up to
110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
contains additional sequences as shown in AAB99132-AAB99143,
CC AAF75903-AAH75920 and AAG66436.
XX
SQ Sequence 655 AA;
```

```
alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent Identity: 22.444
```

## alignment\_block:

US-09-886-400-3 x AAB96088

Align seg 1/1 to: AAB96088 from: 1 to: 655

```

73  GAGAGGATACATCCAGTCCAGTCCAGACACTGATTAAAGAAGAAATTC 122
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
28  GluLysAlaTyrArgProPheLeuGluLeu.....GluGluTyrPr 42
123  T.....TTGGCTCAACATACGGGTATACCTTAAAGTTCCTCC 163
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
42  OasnMetLysValAlaIleHisIleSerGlyLeuValGluTrpLeu 59
164  CGAAG.....GATATTAGACCTCGTTAAAGGGGCATCGCGAGT 204
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
59  GluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75
205  GACCTGATAGATAATCGGACGAGCTACACGACGCACTACTCCCT 254
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
76  GlyGlnValGluValAlaGlyPheTyrGluProValLeuAla 92
255  CTCCTCCGCTAGCAGATAGACGACAGTTTCAGAGAGATAGGAGTTA 304
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
92  aileProLysGluAspArgLeuGluGlnIleTyrLeuLysGluTrpA 109
305  AGGAAGACTCTTCGAGCTTCTCCAAAGGATCTGGCTGCCAGCTC 354
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
109  laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
355  GCCTATGACCCGATATCCCTGCCATACCTGAAGGACACGCTTATGAGTA 404
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
125  ValTrpGlnProGluLeuValLysThrLeuArgGluAlaGlyIleGluTyr 141
405  TCTATCGCCGAGGGGAGGCGATCTTTCTCAGCTCATCTCAACTCG 454
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141  rValValValAsp.....AspTyrHisPheMetSerA 152
455  CG...ATAAGGCAATTAACCGCTCATCCACACCTTATAAGGCGCAA 501
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
152  laGlyLeuSerLysAspGlnLeuPheTrpTyrTyrThrGluAspGly 168
502  AGGGA.....AAGCGCTTAGGTAC..... 522
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
169  GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeuI 185
523  .....ATCAGCTATCTCTCTGCTGTCTCAGG 547
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
185  eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAla 202
548  AGCTTAGGAGCGGATAAGTCGCTTTTGAAGTAAGTAACGCTAAG 597
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
202  erGluAspGluSerLysValAlaValPheHis..... 212
598  GCAGTCAAGACATCGAGCGCTCCGTTTGGTGGCGGTGAACACGGC 647
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
213  .....AspAspGlyGluLysPheGlyIleTrp..... 221
648  TGTATGCTCGGCATCGGAAGCTTCTCTTATGAATCCTAAGAAAGTG 697
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
222  .....ProMet.....Thr 225
698  CGAGCTGGATAGGACAGGAC..... 720
225  yrGluTrpValTyrGluLysGlyTrpLeuArgGluPhePheAspArgVal 241
721  .....AACATTCTTATACGGCACCGCATATAGAG... 750
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242  SerSerAspGluAlaIleAsnIleMetLeuTyrSerGluTyrLeuGlnL 258
751  .....TTCATGGCTATAGGACATTCAGGCTAC..... 780
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
258  sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGlu 275

```

```

781  .....AGATCAGTGTGAGGA 798
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
275  etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
799  TTATTAGAGTTATAGACGAGCTCAACTCGGAAGTGCCTTCCCTCAGA 848
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
291  PheValGluLysLeuLysGluLeuAsn..... 299
849  GCTGAAGCACAGTGAAGGAGCTCTACTTACGGACTTCGAGTTGGGCAC 898
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
300  ....MetPheGluArgTyrArgValPheValArgGlyIleTrp.... 313
899  CAGATAAGAGCTTGAGTATCGAGAGGACGAGGAGGAC..... 939
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
314  ....LysAsnPhe...PheTyrLysTyrProGluAlaAsnTyrMetHis 327
940  GCAAGACTTAATCTCTCTACATAATAGGGGCGAAGTTCGCCCTTTT 989
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
328  LysArgMetLeuMetLeuSer.....ArgLeuLe 337
990  AGCCGAGAACGCGATGCAAGGGATGGGAGCCCTCCTCGAGAGGAGGC 1039
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
337  uArgAspAsnProSerAlaArgPhe..... 346
1040  TGGATGCTTCGGCGCATATATAACGAT.....TGGAGGGGTGAAAAAT 1083
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
347  ....ValLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPhe 361
1084  GGG 1086
|||
362  Gly 362

```

seq\_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:AA47504

seq\_documentation\_block:

ID AA47504 standard; protein; 649 AA.

XX AA47504;

XX 07-JUL-1994 (first entry)

XX Pyrococcus furiosus alpha amylase.

XX Pyrococcus furiosus; alpha amylase; liquefaction; polymers;

XX glucopolymers; thermostable.

XX Pyrococcus furiosus.

XX EP577257-A.

XX 05-JAN-1994.

XX 17-MAY-1993; 93EP-0303801.

XX 09-JUN-1992; 92US-0893928.

XX (UJJO ) UNIV JOHNS HOPKINS.

XX Anfinsen CB, Laderman K;

XX WPI; 1994-009532/02.

XX Purified Pyrococcus furiosus alpha-amylase - used for the

XX industrial liquefaction of gluco-polymers at high temps.

XX Claim 2; Figure 9; 4lpp; English.

XX The purified pyrococcus furiosus alpha amylase can act on substrates  
 CC with a low degree of polymerisation. e.g. glucose polymers as short  
 CC as maltotriose. The enzyme can be used for efficient industrial  
 CC liquefaction of glucopolymers at high temperatures.

XX

SQ Sequence 649 AA;  
 alignment\_scores:  
   Quality: 143.50   Length: 399  
   Ratio: 0.776   Gaps: 24  
   Percent Similarity: 46.366   Percent Identity: 22.807  
 alignment\_block:  
 US-09-886-400-3 x AAR47504 ..  
 Align seg 1/1 to: AAR47504 from: 1 to: 649

```

73 GAGAGGACATACATCCAGTCATCGAGACACTGATTAAAGAGAAATTC 122
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 GlulysCysTyrTrpPheLeuGluThrLeu.....GluGluTyrPr 43
123 T.....TTGGGCTCAACATAAGGGCTATACCTTAAAGTTCTCTCC 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 cAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeu. 59
154 CGAAGGATATT.....ATAGACTCGTTAAAGGGGCATCGG 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 ..GlnAspIleArgProGluTyrIleAspLeuLeuArgSerLeuValLys 75
202 AGTGACCTGATAGATAATCGGAACGAGCTACACGACGCAATACTCC 251
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 ArgGlyGlnValGluIleValAlaIleGlyPheTyrGluProValLeuAl 92
252 CCTCTCCCGCTTAGCAGAGTAGAGCAACAAGTTCAGAGATAGGGAAG 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 aSerIlePro.....LysGluAspArgIleGluGlnIleArgLeuM 106
302 TTAAGGAA.....GAGCTCTTCGAGCTTTCTCAAAGGATTCTGGTG 345
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 eTlysGluTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeu 122
346 CCAAGCTCGCCTATGACCGGATANTCCCTGCCATACCTGAAGACAACGG 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123 ThrGluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGl 139
396 TTATGAGTATCTATCGCGAGGGGAGGCGATCTTTCTCAGCTCATC 445
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 yIleAspTyrValIleValAsp...AspTyrHisPheMetSerAlaGluL 155
446 TCAACTCG.....GCTTTTGAAGCTAAGGTA 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 euSerLysGluLeuLeuTyrTrpProTyrTyrThrGluAspGlyGlu 171
454 .....GGATAAAGCCCAATTAACCCGCTCTATCCACACCTTATAAGGC 497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 ValIleAlaValPheProIleAsp..... 179
498 CCAAGGGAAGAGCGCTTAGGTACATCAGCTATCTCTTGGTCTCAGGG 547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 .....GluLys.....LeuArgTyrLeuIleProPheArgP 190
548 AGCTTAGGAGCGGCAAAAGCTC.....GTTTGAAGCTAAGGTA 588
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 roValAspLysValLeuGluTyrLeuHisSerLeuIleAspGlyAspGlu 206
589 AGCTTAAAGGAGTC.....AAAGCATCGAAGCCGCTACCCGTTTGGGT 632
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 SerLysValAlaValPheHisAspAspGlyGluLysPheGlyIleTrpPr 223
633 GGCCTGTAACAGCGCTGTAATGCTCGGCATCGGAGGCTTCTCTTATGA 682
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 oGlyThrTyrGluTrpValTyr..... 230
683 ATCCTAAGAAAGTGGCGAGCTGGATA..... 708
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 .....GluLysGlyTrpLeuArgGluPheAspArgIleSer 243

```

```

709 ...GAGGACAAGGACAAACATTTCTTATATCGGACCGCATATAGAG..... 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 SerAspGluLysIleAsnLeuMetLeuTyrThrGluTyrLeuGluLysTy 260
751 .....TTCAATTGGCTATAGGACATTGACGCTACAGAATGAGTG 790
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
260 rLysProArgGlyLeuValTyrLeuProIleAlaSerTyr..... 273
791 TTGAGGATTATTAGAGGTTATTAGACGAGCTCAACTCGGAAGTGGCCTT 840
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 .....PheGluMet.....SerGluTrpSerLeu 281
841 CCC.....TCAGAGCTGAAGCA 857
282 ProAlaLysGlnAlaArgLeuPheValGluPheValAsnGluLeuLysVa 298
858 CAGTGA.....AGGAGCTCTACTTACGAGCTTCGAGTTGGG 895
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 lLysGlyIlePheGluLysTyrArgValPheValArgGlyGlyIleTrp. 314
896 CACCAGATAAGAGCTTCAGGATATGGAGAGACGACGAAGGAC..... 939
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
315 .....LysAsnPhe...PheTyr:LysTyrProGluSerAsnTyrMet 327
940 ...GCAAGACTTAAATGCTGTCTCTACAATATAGGGGCGCAACTCGCCT 986
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 HisLysArgMetLeuMetValSerLysLeuValArgAsnAsn..... 341
987 TTTAGCCGAAACAGCGATCCAGGGGATGGGAGCCCTCCCTGAGAGGA 1036
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
342 .....ProGluAla 345
1037 GCGTGGATGCTTCGCGCGATATATAACCAT.....TGGAGGGGT 1077
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 rgLysTyrLeuLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGly 360

```

seq.name: /SDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.AA996091

seq\_documentation\_block:

ID AAB96091 standard; Protein; 1362 AA.

XX AAB96091;

XX 29-OCT-2001 (first entry)

XX Putative P. abyssi amylopullulanase.

XX Hyperthermophilic archaeon; hyperthermophilic protein.

XX Pyrococcus abyssi.

XX FR2792651-Al.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-0005034.

XX 21-APR-1999; 99FR-0005034.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Querellou J, Weissenbach J, Saurin W, Heilig R;

XX WPI; 2001-126236/14.

XX New nucleotide sequences isolated from Pyrococcus abyssi encode

XX proteins useful in industry -

XX Claim 7; Pages 715-719; 1657pp; French.

XX The present invention relates to the genomic sequence of Pyrococcus

CC abyss1 (see AAF86431 and AAH41223-7) and P. abyss1 proteins. P. abyss1 is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyss1 protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO200065062, which  
 CC contains additional sequences as shown in AAB99132-AAB99143,  
 CC AAH75903-AAH75920 and AAG66436.  
 XX  
 SQ Sequence 1362 AA;

## alignment\_scores:

Quality: 124.00 Length: 381  
 Ratio: 0.756 Gaps: 16  
 Percent Similarity: 43.045 Percent Identity: 21.522

## alignment\_block:

US-09-886-400-3 x AAB96091 ..

Align seg 1/1 to: AAB96091 from: 1 to: 1362

```

94 ATCGAGACACTGATTAAGAGAAATTCCTTTGGGCTCAACATAACGGG 143
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
236 ValGluThrValLeuLysHisGlnMet...TrpLeuLeuAsnHisThr.. 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
144 CTATACCTTAAGTCTCCCGAGGATATATAGACCTCGTTAAAGGGG 193
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 .....PheGluGluHisGluLysIleAsnLeuLeuGly. 262
194 GCATCGGAGTGACTCTAGAGATAATCGAAGAGCTACACGCGCA 243
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 .....AsnGlyAsnValGluValThrValProTyrThrHisPro 276
244 ATACTCCCTCCCTCCCG.....CTTAGCAGAGTAGAGCACA 281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
277 IleGlyProIleLeuAsnAspPheGlyTrpTyrGluAspPheAspAlaG 293
282 AGTCAGAGAGATAGGGAAGTTAAGGAAGAGCTCTTC.....GAGC 322
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
293 nValLysLysAlaAsnGluLeuTyrLysGluTyrLeuGlyAlaGlyLysV 310
323 TTCTCCAAAGGATCTGGCTGCACAGCTCGCTATGACCCGATAATC 372
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 alThrProLysGlyGlyTrpAlaAlaGluSerAlaLeuAsnAspLysThr 326
373 CCTGCATACCTGAAGCAACAGGTTATGAGTATCTATTCGCGACGGG.. 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
327 LeuGluIleLeuAlaGluAsnGlyTrpLysTrpValMetThrAspGlnLe 343
421 .....GAGCGGATGCTTTCTCAGCTCATCTCACTCGCGGATAAAGC 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
343 uValLeuGluLysLeuGlyValProLysThrIleGluSerTyrTyrLysP 360
464 CA..... 465
360 rotPrpValAlaGlnPheGlyAspLysLysIleTyrLeuPheProArgAsn 376
465 ..... 465
377 HisAspLeuSerAspArgValGlyPheArgTyrAlaGlyMetAsnGlnTy 393
466 .....ATTAACCGCTCTATCCACACCTTATAAGGCCCAAGGGAAA 508
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
393 rAspAlaValLysAsnPheValGluGluLeuLysIleGlnLysGlnA 410
509 ACCGCTTTAGGTACATCAGCTATCTCTGCTCTCAGGAGCTTAGGAAG 558
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 snTyrAspGlySerLeuValTyrValIleThrLeuAspGlyGluAsnPro 426
559 GCGATAAGCTGTTTTGAAGTAAGTACGCTAAAGGCAGTC...AA 605
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

427 TrpGluHisTyrProPheAspGlyLysLeuPheLeuGluLeuTyrTr 443
606 AGACATCGAAGCCGCTACCCGTTGGTGGCGGTGAACACGGCTGTA... 651
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
443 gGlnLeuGluLeuLeuLysLysGlyLeuIleArgThrValThrProS 460
652 .....ATCCTCGGC.....ATCGAAGGCTTCCTCTATG 681
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
460 erGluTyrIleGluMetPheGlyAspLysAlaAsnLysLeuThrProLys 476
682 AATCCTAAGAAAGTGGCGAGCTGGATAGAGGACAGCAACACTTCTCT 731
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
477 MetMetLysArgLeuAspPheThrThrGluAspAsnValAsnAlaLeu 493
732 ATACGGCACCGATATAGAGTTCATGGCTATAGGACATTGCAGGCTACA 781
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
493 uLysAlaLysThrLeu.....GlyGluL 501
782 GAATGAGTGTGAGGATTATTAGAGCTTATAGAGCTTATAGAGCTCA 831
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 euTyrAspMetValGlyValThrGlu..... 509
832 CTGTCCTTCCTCAGAGCTGAACACAGTGGAGGAGGAGCTCTACTACG 881
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
510 .....GluMetGlnTrpPr 514
882 GACTCGAGTGGGCACAGATAAGAGCTTGAGATATGGAGAGAGGACG 931
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
514 oGluSerSerTrpIle...AspGlyThrLeuSerThrTrp.....I 527
932 AAGGACGCAAGACTTAATATGCTCTACATATATGAGGGGCGAATC 981
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
527 leGlyGluProGlnGluAsnIleAlaTrpTyrTrpLeuTyrLeuAla 543
982 CCCTCTTTAGCCGAGAACAGCAT...GCAAGGGATGGAGGCCCTCC 1028
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
544 LysAlaLeuPheGluAsnLysAspAsnValLysAspTrpAsnLysAla 560
1029 TGAGAGGAGCTGGATGCTTCGCGGCGATATATACGATGG 1071
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
560 rGluTyr.....LeuPheArgAlaGluGlySerAspTrp 571

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:AAW54870

## seq\_documentation\_block:

ID AAW54870 standard; Protein; 653 AA.

XX AAW54870;

XX AC

XX XX

XX 01-SEP-1998 (first entry)

XX DE Super heat resistant 4-alpha-glucanotransferase.

XX KW Super heat-resistant 4-alpha-glucanotransferase; heat-treatment;

XX KW alpha-1,4-glucan; alpha-1,4-glucoside bond.

XX OS Pyrococcus sp.

XX PN JP10150986-A.

XX XX 09-JUN-1998.

XX XX 21-NOV-1996; 96JP-0311117.

XX XX 21-NOV-1996; 96JP-0311117.

XX PA (BEAB-) BE ABLE KK.

XX PA (NAGA-) NAGASE SEIKAGAKU KOGYO KK.

XX XX WPI; 1998-379989/33.

XX DR N-PSDB; AAV27026.

XX PT New 4-alpha-glucanotransferase which has been heat-treated - used at



PT high temperatures to transfer at least one glucose unit

XX Claim 4; Page 10-13; 18pp; Japanese.

CC The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of  
 CC 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at  
 CC least 90% activity after heat-treatment at 100 deg. C for 30 minutes at  
 CC pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan  
 CC to alpha-1,4-glucan by an alpha-1,4-glucoside bond.

XX Sequence 653 AA;

alignment\_scores:  
 Quality: 121.50 Length: 386  
 Ratio: 0.63 Gaps: 21  
 Percent Similarity: 48.187 Percent Identity: 21.244

alignment\_block:

US-09-886-400-3 x AAW54870 ..

Align seg 1/1 to: AAW54870 from: 1 to: 653

```

73 GAGAGGATACATCCAGCATCTGAGACACTGATTAAAGAAATTC 122
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
28 GluArgSerTyrArgProPheMetGluThrLeu.....GluGlyTrp 42
123 T.....TTTGGCTCAACATAACGGCTATACCTTAAAGTTCCTC. 162
|  |||  |||  |||  |||  |||  |||  |||  |||  |||
42 OasnMetLysValAlaValHisTyrSerGlyProLeuLeuGluTrpIleA 59
163 .....CCGAGGATATATAGACTCTCTTAAGGGGCGCGAGT 204
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
59 rgAspAsnLysProGluHisLeuAspLeuLeuArgSerLeuValLysArg 75
205 GACCTGTAGAGATAATCGGAACGAGCTACACGACGACATACCTCCCT 254
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
76 GlyGlnLeuGluIleValAlaGlyPheTyrGlyProValLeuAla 92
255 CCTCCGCTTAGCAGATAGACGACAAAGTTCAGAGATAGGGAAGTTA 304
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
92 rIleProGluAspArgIleValGlnIleGluLysLeuLysGluPheA 109
305 AGGAAGAGCTCTTCGAGCTTCTCCAAAGGATTCTGGTCCGACAGCTC 354
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
109 laArgAsnLeu...GlyTyrGluAlaArgGlyValTrpLeuThrGluArg 124
355 GCCTATGACCGATAATCCCTGCCATCTAGAGGACAAACGGTTATGAGTA 404
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
125 ValTrpGlnProGluLeuValLysSerLeuArgAlaAlaGlyIleAspTy 141
405 TCTATTCGCGACGGGAGCGGATGCTTTCTCAGCTCATCTCAACTCGG 454
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
141 rValIleValAsp.....AspTyrHisPheMetSerA 152
455 CG...ATAAGCCCAATTAACCGCTCTATCCACACCTTATAAGGCCCAA 501
||  |||  |||  |||  |||  |||  |||  |||  |||  |||
152 laGlyLeuSerLysAspGluLeuPheTrpProTyrTrpThrGluAspGly 168
502 AGGGA.....AAGCGTTTGGTACATCAG 527
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArg..... 182
528 CTATCTCCTTGGTCTAGGAGGCTTAGGAAGGCGATAAAGCTCGTTT. 576
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
183 TyrLeuIleProPheArgProValAspLysThrLeuGluTyrLeuHis 199
577 .....GAGGTAAAGTAAACGTAAGGCAGTC.....AAAGACATC 612
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
199 erLeuAspAspGlyAspGluSerLysValAlaValPheHisAspAspGly 215
613 CAAGCGGTACCGTTTGGGTGGCGGTGAACACGGCTGTAATGCTGGCAT 662
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

216 GluLysPheGlyValTrpProGlyThrTyrGluTrpValTyr..... 229
663 CGGAAGGCTTCTCTTATGAATCTAAGAAAGTGGGAGCTGGATA.... 708
230 .....GluLysGlyTrpLeuArg 236
709 .....GAGCACAGACACATCTTCTTATAC 735
236 luPhePheAspArgValSerSerAspGluArgIleAsnLeuMetLeuTyr 252
736 GGCACCGATATAGAG.....TTCAATTGGCTATAGGGGACAT 770
253 SerGluTyrLeuGlnArgPheArgProArgGlyLeuValTyrLeuProIl 269
771 TGCAGGCTAC...AGAATGAGTGTGAGGGATTA..... 801
269 eAlaSerTyrPheGluMetSerGluTrpSerLeuProAlaArgGlnAlaL 286
802 .....TTAGAGTTATAGAGGAGCTCAACTCGGAAGTGTGCCTTCCC 843
286 ysLeuPheValGluPheValGluLeuLys..... 296
844 TCAGAGCTGAAGCACAGCTGGAAGGAGCTCTACTTACGAGCTTCGAGTTG 893
297 LysGluAsnLysPheAspArgTyrArgValPheValArgGlyIleTr 313
894 GGCACACAGATAAGAGCTTGAGGATATGGAGAGGACGACGAAGGAAC... 939
313 p.....LysAsnPhe...PhePheLysTyrProGluSerAsnTyrM 326
940 .....GCAAGACTTATATCTCTCTACAAATATAGGGGCGGAACTCGCC 984
326 eHisLysArgMetLeuMetValSerLysAlaValArgAsnAsn..... 340
985 CTTTTCGCGAGAACAGCGATGCAAGGGGATGGAGCCCTCCCTCGAGAG 1034
341 .....ProGluAla 343
1035 GAGCTGGATGCTTCGGCGGATATATAACGAT.....TGGAGGGGTG 1078
343 aArgGluPheIleLeuArgAlaGlnCysAsnAspAlaTyrTrpHisGlyV 360
1079 AAAATGGG 1086
360 alPheGly 362
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.AAG82528
seq_documentation_block:
ID AAG82528 standard; Protein; 360 aa.
XX
AC AAG82528;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 990S-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;

```

XX WPI; 2001-316495/33.  
DR N-PSDB; AAH53378.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 578-579; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG81120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 360 AA;

alignment\_scores:

Quality: 104.50 Length: 299  
Ratio: 0.731 Gaps: 12  
Percent Similarity: 47.826 Percent identity: 19.398

alignment\_block:

US-09-886-400-3 x AAG82528

Align seg 1/1 to: AAG82528 from: 1 to: 360

10 CTCGCTTTTCAGCGCAACCTCCAGTATCCGAAATCCCAAGAGCGAAAT 59  
||||| |||  
61 LeuilePheLysGly.....Va 66  
60 CCCAAGGTCATAGAGAGCGATATCCAGTCATCGAGACATGATTA 109  
: ||||| ||| : |||||  
66 lLysArgIleValGluAspGlyTyrGlySerIleIleArgLysLeuileG 83  
: ||||| ||| : |||||  
110 AAGAAGAAATTCCTTTGGGCTCAACATAACGGGCTATACCTTAAAGTTC 159  
: ||||| ||| : |||||  
83 llnAsnAsnle.....AsnLeuileAlaLeuHisThrAsnLeuAspVal 97  
160 CTCGCCAAGGATATATAGACCTCTTAAAGGGGCATCGCGAGTGACCT 209  
||||| : ||||| : |||||  
98 AsnProLysGlyValAsnArgMetLeuAlaAspGlnIleGlyLeuGluAs 114  
210 GATAGAGATAATCGGAACG.....AGCTACACGACGCAATACCTCCGCC 253  
||||| : ||||| : |||||  
114 nlieSerMetIleAsnThrAsnSerTyrTyrTyrLysValGlnThrP 131  
254 TCTCCCGGTTAGCAGAGTAGAGCACACAGTTTCAGAGAGATAGGGAAGTT 303  
: ||||| : ||||| : |||||  
131 heileProLysAsnTyrIleGlu.....AspPhe 140  
304 AAGAGAGAGCTCTTCAGGCTTTCTCCAAAGGATTCGGCTCCAGAGCT 353  
||||| : ||||| : |||||  
141 LysAspSerLeuAsnGluLeu..... 147  
354 CGCCTATGACCCGATAATCCCTGCCATCTGAAGGACACGCTTATGAGT 403  
: ||||| : ||||| : |||||

148 .....GlyLeuAlaLysGluGlyAsnTyrGluT 157  
404 ATCTATTCCGCGAGCGGAGCGGATGCTTTTCTCAGCTCATCTCAACTCG 453  
||| : ||||| : |||||  
157 yTCysPhePheGluSerGluGly.....LysGly 166  
454 GCGATAAGCGCAATTAACCGCTCTATCCACACCTTATAAGGCCCAAG 503  
: ||||| : ||||| : |||||  
167 GlnPheLysProValGlyAspAlaSerProTyrIleGlyLysLeuAspSe 183  
504 GGAAGAGCGCTTAGGTATCATCATGCTATCTCTTGGTCTCAGGAGCTTA 553  
183 r.....lleglTyrVal..... 187  
554 GGAAGCGGATAAGCTCTGTTTT.....GAAGTAAAGGTA 588  
: ||||| : ||||| : |||||  
188 ..AspGluIleLysLeuGluPheMetIleLysAspAsnGluLeuGluile 203  
589 AGCTAAAGGAGTCAAGAGCATCGAA.....GCCGTACCGCTT..... 627  
204 ThrLysArgAlaIleLeuAspAsnHisProTyrGluThrProValPheAs 220  
628 .TGGTGGCCGTGAACACGGCTGTAACTCGGCATCGGAAGGCTTCTC 676  
: ||||| : ||||| : |||||  
220 pPheIleLysMetAsnLysGluSerGluTyrGlyLeuGlyIleIleGlyG 237  
677 TTATGATCTTAAGAAAGTGGCGAGCTGGATAGGAGGACAGGACAACTT 726  
: ||||| : ||||| : |||||  
237 llnLeuAsnGlnThrMetThrLeuAspGluPheSerGluTyrAlaLysLys 253  
727 CTTCTATACGGCAGCATATAGATTTCATGCTATAGGACATTCGAGG 776  
||| : ||||| : |||||  
254 GlnLeuAsnIleProSerValArgTyrThrGlyGlnHisAspSerProIl 270  
777 CTACAGATGAGTGTAGGGGATTTATAGAGTT.....A 811  
: ||||| : ||||| : |||||  
270 eLysLysValAlaIleIleGlyGlySerGlyIleGlyPheGluTyrLysA 287  
812 TAGACGAGCTCAACTCGGAAGTGGCTTCCTCCTCAGAGCTGAAGCAC 858  
: ||||| : ||||| : |||||  
287 lAsrGlnLeuGlyAlaAspValPheValThrGlyAspIleLysHis 302  
seq\_name: /SIBSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:AA82506  
seq\_documentation\_block:  
ID AAG82506 standard; Protein: 376 AA.  
XX  
AC AAG82506;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2106.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
PI Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
DR N-PSDB; AAH53356.  
XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 569; 2188pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 376 AA;

alignment\_scores:  
Quality: 104.50 Length: 299  
Ratio: 0.731 Gaps: 12  
Percent Similarity: 47.826 Percent Identity: 19.398  
alignment\_block:  
US-09-886-400-3 x AAG82506 ..  
Align seg 1/1 to: AAG82506 from: 1 to: 376

10 CTCCTCTTTCACGCACCTCCAGTATGCCGAATCCCAAGAGCGCAAT 59  
||||:|||||  
47 LeuilePheLysGly.....Va 52  
50 CCCAAGGTCATAGAGGACATACATCCAGTCATCGACACATGATTA 109  
:||||:|||||  
52 lLysArgileValGluAspGlyTyrGlySerIleIleArgLysLeuileG 69  
:||||:|||||  
110 AAGAGAAATCTTTGGGCTCAACATACGGGCTATACCTAAAGTTC 159  
:||||:|||||  
69 InAsnAsnIle.....AsnLeuIleAlaLeuHistrAsnLeuAspVal 83  
|||||  
160 CTCCTCCGAGGATATTATAGACCTCGTTAAAGGGGCGATCGAGTCACCT 209  
|||||  
84 AsnProLysGlyValAsnArgMetLeuAlaAspGlnIleGlyLeuGluAs 100  
210 GATGAGATATCCGAAGC.....AGCTACAGCGACGCAATACCTCCCC 253  
||||:|||||  
100 nileSerMetIleAsnThrAsnSerSerTyrTyrTyrValGlnThrP 117  
:||||:|||||  
254 TCCCTCCGCTTAGCAGAGTACAGACACAACTTCAGAGATAGGAGTT 303  
:||||:|||||  
117 heileProLysAsnTyrIleGlu.....AspPhe 126  
304 AAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTCCTGGCTGCAGAGCT 353  
||||:|||||  
127 LysAspSerLeuAsnGluLeu..... 133  
354 GCGCTATGACCCGATATCCCTGCCATCTGAAGGACACGGTTATGAGT 403  
||||:|||||  
134 .....GlyLeuAlaLysGluGlyAsnTyrGlnT 143  
404 ATCTATTCCGCGAGGGGAGCGAGTCTTTTCAGCTCATCTCAACTCG 453  
|| ||| :||||:||||:

143 yrCysPhePheGluSerGluGly.....LysGly 152  
454 GCGATAAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCAAG 503  
:||||:|||||  
153 GlnPheLysProValGlyAspAlaSerProTyrIleGlyLysLeuAspSe 169  
504 GGAAGACCGCTTTAGGTACATCAGCTATCTCCTTGCTCTCAGGAGCTTA 553  
||||:|||||  
169 r.....IleGluTyrVal..... 173  
554 GGAAGCGGATAAAGCTCGTWTTT.....GAAGTAAGGTA 588  
||||:|||||  
174 ..AspGluIleLysLeuGluPheMetIleLysAspAsnGluLeuGluile 189  
589 ACGTAAAGCAGCTCAAGACATCGAA.....GCCGTACCCGTT..... 627  
||||:|||||  
190 ThrLysArgAlaIleLeuAspAsnHisProTyrGluThrProValPheAs 206  
628 .TGCGTGGCGCTGAACAGGCTGTATGTCTGGCATCGGAAGCTTCCTC 676  
:||||:|||||  
206 pPheIleLysMetAsnLysGluSerGluTyrGlyLeuGlyIleIleGlyG 223  
677 TTATGAATCCTAAGAAAGTGGCGAGCTGGATAGGACAGGACACACATT 726  
:||||:|||||  
223 InLeuAsnGlnThrMetThrLeuAspLupheserGluTyrAlaLysLys 239  
727 CTCCTATACGGCACCATATAGAGTTTCATGTCTAGGACATTCGACAG 776  
||||:|||||  
240 GlnLeuAsnIleProSerValArgTyrThrGlyGlnHisAspSerProI 256  
777 CTACAGATGAGTGTTCAGGAGATTATAGAGTT.....A 811  
256 eLysLysValAlaIleIleGlyLysSerGlyIleGlyPheGluTyrLysA 273  
812 TAGACGAGCTCAACTCGAACTGTGCTTCCTCCCTCAGAGCTGAAGCAC 858  
:||||:|||||  
273 laSerGlnLeuGlyAlaAspValPheValThrGlyAspIleLysHis 288  
seq.name: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.AAR94013  
seq\_documentation\_block:  
ID AAR94013 standard; Protein; 329 AA.  
XX  
AC AAR94013;  
XX  
DT 21-AUG-1996 (first entry)  
XX  
DE Heat resistant maleate dehydrogenase.  
XX  
KW Heat resistant maleate dehydrogenase; h-rMAD; NADH; L-aspartic acid;  
KW alpha-ketoglutaric acid; glutamine oxalo-transaminase activity; GOT.  
XX  
OS Bacillus stearothermophilus ATCC 12016.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 185  
FT /note= "Given in the specification as Var"  
XX  
PN JP08047389-A.  
XX  
PD 20-FEB-1996.  
XX  
PF 01-JUL-1994; 94JP-0151045.  
XX  
PR 03-JUN-1994; 94JP-0121629.  
XX  
PR 02-JUL-1993; 93JP-0164701.  
XX  
XX (TOYM ) TOYOBO KK.  
XX WPI; 1996-166248/17.  
XX DR N-PSDB; AAT17715.  
XX Protein having heat resistant maleate dehydrogenase activity - and  
PT



```
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145215.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154778.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 104.00 Length: 365
Ratio: 0.838 caps: 20
Percent Similarity: 44.658 Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG16038 ..
Align seg 1/1 to: AAG16038 from: 1 to: 571

85 ATCCAGTCATCGACACTGATTAAGAAGAAATCTTTTGGGCTCAA 134
101 LeuProileLeuGlu...MetValProGluGluLeuHisPhe..... 113
135 CATACACGGGCTATACCTAAAGTTCTCCCGAAGGATATTATAGACCTCG 184
114 ThrValProTyrAlaLeuSerPheileMetGlyAspProileLysMetA 130
185 TTAAGGGGCGATCGGACGTGACCTG.....ATAGAGATA 219
130 lathrLeuGlyIleAspAsnGlnLeuProThrGlyVallysileGlyLys 146
```



PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139829.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 25-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## alignment\_scores:

Quality: 104.00 Length: 365  
Ratio: 0.638 Gaps: 20  
Percent Similarity: 44.658 Percent Identity: 22.466

## alignment\_block:

US-09-886-400-3 x AAG48655 ..

Align seg 1/1 to: AAG48655 from: 1 to: 571

85 ATCCGAGTCATCGAGACACTGATTAAGAGAAATTCCTTTGGGTCAA 134

```

101 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 113
135 CATACGGGCTATACCTTAAGATTCTCCCGAAGATATATACACCTCG 184
114 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMet 130
185 TTAAGGGGGCATCCGAGTACCTG.....ATAGAGATA 219
130 laThLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 146
220 ATCGGAACAGAGTACACGACGCAATATCCCTCCCTCCGCTAGCAG 269
147 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerG1 161
270 AGTAGAAGCACAAAGTTCAGAGAGAT..... 294
161 uLeuGlyGlyIleIleProArgGluThrLeuLeuTrpLysLeuLysLeu 178
295 .....AGGGAAGTTAAGGAA 309
178 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 194
310 GAGCTCTTCGAGCTT...TCTCCAAGGATTCGTGGCTGCCAGAGCTCGC 356
195 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnG1 211
357 CPTAGACCGGATAATCCCTGCCATACTG..... 384
211 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysp 228
385 ..AAGGACACAGGTTATGAGTATCATTCGCGGAGGGGAGCGGATGCTT 432
228 helyAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 244
433 TTCTCAGCTCATCTCACTCGCGGATAAAGCAATTAACCGCTCTATCC 482
245 .....ThrValIleLysGlyThrGly..... 251
483 ACACCTTATAAGGCCCAAGGAAACCGCTTTAGGTACATCAGC..... 528
252 .....LysTyrArgArgSerTrpArgTyrAspLeuValSerAsp 265
529 .....TATCTCCTTGGT.....CTC 543
265 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 281
544 AGGAGCTTAGGAGCGGATAAAGCTCGTTTT.....GA 578
282 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 298
579 AGGTAGGTAAAGCTAAAGCCAGTCAACACATCGAACCCCTACCC..... 624
298 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 311
625 .....GTTTGGTGGCGCGTGAACAGCGTGTAAATCCTCGGCTAGGAAG 669
311 yselyProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 327
670 CTTCTCTTATGATCCTTAAGAAAGTGGCGAGCTGGATAGAGGACAGGA 719
328 GlyProMetSerGlu.....AlaPheIleLysGluLys... 338
720 CAACATCTCTCTA..... 732
339 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 355
733 .....TACGGCAGCATATAGATTTCATTGCTTAT 762
355 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 370
763 AGGACATTCAGCGCTACAGATGAGTGTGAGGGATTATAGAGGTTAT 812

```

```

371 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 383
813 AGACGAGCTCAACTCGGAACCTGCCTTCCCTCA.....GAGCTGA 853
383 uaspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 400
854 AGCACAGTGGAGGAGCTCTACTACGAGCTTCAGTTGGGACCAGAT 903
400 euHisAsnArgGlyGluGlnTyr..... 407
904 AAGACCTTGAGCATATGAGAGAGAGGACGACGACGACGACACTT 948
408 .....LysLeuIleIrrProGluGlnGlnGluPheValArgMet 420
seq_name: /SIBS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AA616037
seq_documentation_block:
ID: AA616037 standard; Protein; 704 AA.
XX
AC: AAG16037;
XX
DT: 17-OCT-2000 (first entry)
XX
DE: Arabidopsis thaliana protein fragment SEQ ID NO: 16523.
XX
KW: Protein identification; signal transduction pathway; metabolic pathway;
KW: hybridisation assay; genetic mapping; gene expression control; promoter;
KW: termination sequence.
XX
OS: Arabidopsis thaliana.
XX
PN: EP1033405-A2.
XX
PD: 06-SEP-2000.
XX
PF: 25-FEB-2000; 2000EP-0301439.
XX
PR: 25-FEB-1999; 99US-0121825.
PR: 05-MAR-1999; 99US-0123180.
PR: 09-MAR-1999; 99US-0123548.
PR: 23-MAR-1999; 99US-0125788.
PR: 25-MAR-1999; 99US-0126264.
PR: 29-MAR-1999; 99US-0126785.
PR: 01-APR-1999; 99US-0127462.
PR: 06-APR-1999; 99US-0128234.
PR: 08-APR-1999; 99US-0128714.
PR: 16-APR-1999; 99US-0129845.
PR: 19-APR-1999; 99US-0130077.
PR: 21-APR-1999; 99US-0130449.
PR: 23-APR-1999; 99US-0130510.
PR: 28-APR-1999; 99US-0130891.
PR: 30-APR-1999; 99US-0131449.
PR: 30-APR-1999; 99US-0132048.
PR: 04-MAY-1999; 99US-0132484.
PR: 05-MAY-1999; 99US-0132485.
PR: 06-MAY-1999; 99US-0132486.
PR: 07-MAY-1999; 99US-0132863.
PR: 11-MAY-1999; 99US-0134256.
PR: 14-MAY-1999; 99US-0134218.
PR: 14-MAY-1999; 99US-0134219.
PR: 14-MAY-1999; 99US-0134221.
PR: 14-MAY-1999; 99US-0134370.
PR: 18-MAY-1999; 99US-0134768.
PR: 19-MAY-1999; 99US-0134941.
PR: 20-MAY-1999; 99US-0135124.
PR: 21-MAY-1999; 99US-0135353.
PR: 24-MAY-1999; 99US-0135629.
PR: 25-MAY-1999; 99US-0136021.
PR: 27-MAY-1999; 99US-0136392.
PR: 28-MAY-1999; 99US-0136782.
PR: 01-JUN-1999; 99US-0137222.

```



PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452..  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145216.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148585.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156595.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:

Quality: 104.00 Length: 365  
 Ratio: 0.538 Gaps: 20  
 Percent Similarity: 44.558 Percent Identity: 22.466

alignment\_block:  
 US-09-886-400-3 x AAG16037 ..  
 Align seg 1/1 to: AAG16037 from: 1 to: 704

```

85 ATCCAGTCATCGACACTGATTAAGACGAATTCCTTTGGCTCAA 134
234 LeuProIleLeuGlu...MetValProGluLeuHisPhe..... 246
135 CATAACGGGCTATACCTAAAGTCTCTCCGAGGATATTAGACCTCG 184
247 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMet 263
185 TTAAGGGGGATCCGGAGTACCTG.....ATAGAGATA 219
263 laThrLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 279
220 ATCGGAACGAGCTACGACGACCAATCTCCCTCTCCGCTAGCAG 269
280 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerG 294
270 AGTAGAAGCAAGTTCAGAGAGAT..... 294
294 uLeuGlyGlyIleIleProArgGluThrLeuLeuThrLysLeuLysLeu 311
295 .....AGGGAAGTTAAGGAA 309
311 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 327
310 GAGCTCTTCGAGCTT...TCTCCAAGGATTCCTGGTCCGAGAGCTCGC 356
328 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnG 344
357 CTATGACCGGATAAATCCCTGCCATCTG..... 384
344 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 361
385 ..AAGGACACGCTTATGATGATCTATTCGCGACGGGAGCGGATGCTT 432
361 heLysAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 377
433 TTCTCAGCTCATCTCAACTCGCGGATTAAGCAATTAACCGCTCTATCC 482
378 .....ThrValIleLysGlyThrGly..... 384
483 ACACCTTTATAAGGCCCAAGGGAAGGCGCTTTAGGTACATCAGC.... 528
385 .....LysTyrArgArgSerTrpArgTyrAspLeuValSerAspP 398
529 .....TATCTCCTTGGT.....CTC 543
398 heLeuProProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 414
544 AGGAGCTTAGGAAGCGGATAAAGCTCGTTT.....GA 578
415 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 431
579 AGGTAGGTAACTGCTAAAGCAGCTCAAGACATCGAAGCCGTACCC.... 624
431 pGlyLysIle.....ValLysGlyLeuAlaGlyValProAspL 444
625 .....GTTTGGTGGCGGTGAACACCGCTGTATGCTCGCATCGGAAGG 669
444 ysGlyProValLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 460
670 CTTCCTCTTATGAATCCTAAGAAGTGGGAGCTGGATGAGACACAAGGA 719
461 GlyProMetSerGlu.....AlaPheIleLysGluLys... 471
  
```

```

720 CAACATCTTCTTA..... 732
472 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 488
733 .....TACGGCACCGATATAGAGTTTCATTTGGCTAT 762
488 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 503
763 AGGACATTCGAGCTACAGATAGTGTGAGGGATATTAGAGTTAT 812
504 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 516
813 AGACAGCTCACTCGGAACCTGTCCTCCCTCA.....GAGCTGA 853
516 uaspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 533
854 AGCACAGTGGAGGAGGAGCTCTACTTACGAGCTTGGAGTTGGCACCAGAT 903
533 euHisAsnArgGlyGluGlnTyr..... 540
904 AAGAGCTTGAGGATATGGAGAGAGGACGACGGAACGCAAGACTT 948
541 .....LysLeuIleTrpProGlnGlnGluPheValArgMet 553
seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:AAAG48654
seq_documentation_block:
ID AAG48654 standard; Protein: 704 AA.
XX
AC AAG48654;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
  
```

PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138450.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 26-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.



PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134376.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 15-JUL-1999; 99US-0144086.  
PR 15-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149830.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160817.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161040.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 104.00      Length: 365
  Ratio: 0.638         Gaps: 20
  Percent Similarity: 44.638  Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG16036 ..
Align seg 1/1 to: AAG16036 from: 1 to: 726

85 ATCCAGCTATCGACACTGATTAAGAAGAAATCTCTTTGGGCTCAA 134
256 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 268
135 CATAAGGGCTATACCTTAAAGTCTCCCGAAGGATATTATAGACPCG 184
269 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLeuMet 285
185 TTAAGGGGCTATCGAGTACGCTGACCTG.....ATAGAGATA 219
285 IaThrLeuGlyIleAspAnGlnLeuProThrGlyValIleGluLys 301
220 ATCGGAACGAGCTACAGCACGCACTATCCCTCCCTCCGCTAGCAG 269
302 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 316
270 AGTAGAAGCACAAAGTTCAGAGAGAT..... 294
316 uLeuGlyGlyIleIleProArgGluThrLeuLeuTrpLysLeuLysLeuL 333
295 .....AGGGAAGTTAAGGAA 309
333 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 349
310 GAGCTCTTCGAGCTT...TCTCCAAAGGATCTGGCTGCCAGAGCTCGC 356
350 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnGI 366
357 CTATGACCGGATAATCCCTGCCACTG..... 384
366 uGluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 383
385 ..AAGCACACGGTTATAGTATCTATTCGCCGACGGGAGCGGATGCTT 432
383 helysAspAnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 399
433 TTCTGAGCTATCTCAACTCGGCGATAAGCAATTAACCGCTCTATCC 482
400 .....ThrValIleLysGlyThrGly..... 406
483 ACACCTTATAAGCCCAAGGAAAGCGCTTAGGTACATCAGC..... 528
407 .....LysTyrArgSerTrpArgTyrArgTrpArgTrpArgTrpArgTrp 420

```

```

529 .....TATCTCCTGGT.....CTC 543
420 heLeuProSerLysGlyGluLeuAlaTyrAlaLeuAspGluValLeu 436
544 AGGAGCTTAGCAAGCGCATAAAGCTCGTTTT.....GA 578
437 GlyPheLeuArgAsnAlaValGlySerValPheSerThrMetGluAs 453
579 AGTAAGGTAAAGCTAAAGGAGTCAAAAGCATCAAGCCGTACCC.... 624
453 polyLysile.....ValLysGlyLeuAlaGlyValProAspL 466
625 .....GTTTGGTGCCCGTGAACAGCGCTGTAATGCTCGCATCGGAAG 669
466 ysglyProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 482
670 CTCTCTTATGAATCTTAAGAAAGTGGCGAGCTGGATAGAGGACAAGGA 719
483 GlyProMetSerGlu.....AlaPheIleLysGluLys... 493
720 CAACATTCTTCTA..... 732
494 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 510
733 .....TACGGCACCGATATAGAGTTTCATTTGGCTAT 762
510 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 525
763 AGGGACATTGCAGGCTACAGATGAGTGTGAGGATATTAGAGTTAT 812
526 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 538
813 AGACGAGCTCAACTCGGAACGTGCTTCCCTCA.....GAGCTGA 853
538 uaspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 555
854 ACACAGCTGGAAGGAGCTCTACTTACGAGCTTGGAGTTGGCACCAGAT 903
555 euHisAsnArgGlyGluGlnTyr..... 562
904 AAGAGCTTGAGGATATGGAGAGACGACGAAGGACGCAAGACTT 948
563 .....LysLeuIleTrpProGluGlnGluPheValArgMet 575

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: AAG48653
seq_documentation_block:
ID AAG48653 standard; Protein; 726 AA.
XX
AC AAG48653;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61463.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.

```

PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 15-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 15-JUN-1999; 99US-0139452.  
PR 15-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141642.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144326.  
PR 19-JUL-1999; 99US-0144327.  
PR 19-JUL-1999; 99US-0144328.  
PR 19-JUL-1999; 99US-0144329.  
PR 19-JUL-1999; 99US-0144330.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 29-SEP-1999; 99US-0156458.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.

```

PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159368.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 104.00      Length: 365
  Ratio: 0.638         Gaps: 20
  Percent Similarity: 44.658  Percent Identity: 22.466

alignment_block:
US-09-886-400-3 x AAG48653
..
Align seg 1/1 to: AAG48653 from: 1 to: 726

85 ATCCAGTCATCGACACATGATTAAAGAAATCTTTTGGGCTCAA 134
   :::::::::::::: ::::: :::::::::::::: :::::
256 LeuProIleLeuGlu...MetValProGluGluLeuHisPhe..... 268

135 CATACGGGGTATACCTTAAGTTTCCTCCGAGAGGATATTATAGACTCG 184
   :: :::::::::::::: ::::: :::::::::::::: :::::
269 ThrValProTyrAlaLeuSerPheIleMetGlyAspProIleLysMetA 285

185 TTAAGGGGCGATCGCGAGTGACCTG.....ATAGACATA 219
   :::::::::::::: ::::: :::::::::::::: :::::
285 IatThrLeuGlyIleAspAsnGlnLeuProThrGlyValLysIleGluLys 301

220 ATCGSACGAGCTACACGACGCAATACTCCCTCCCTCCGCTTAGCAG 269
   :: :::::::::::::: ::::: :::::::::::::: :::::
302 LeuArgGlnArgLeuThrLysThrMetLeuProLeu.....LeuSerGI 316

270 AGTAGAACCAACAAGTTCAGACAGAT..... 294
   ::::: ::::: ::::: :::::
316 uLeuGlyIleIleProArgGluThrLeuLeuTrpLysLeuLysLeuL 333

295 .....AGGGAAGTTAAGAA 309
   :::::
333 euArgSerGlyCysAlaTyrAlaAsnSerArgIleHisAlaValGlnAla 349

310 GAGCTCTTCGAGCTT...TCTCCAAAGGATTCTGGCTGCCAGAGCTCCG 356
   ::::::::::: ::::: :::::
350 GluValLeuValLeuAlaSerGlyLysAspMetMetLeuProSerGlnGI 366

357 CTATGACCCGATATCCCTGCCATACG..... 384
   :::::
366 uCluAlaLysArgLeuHisGlyLeuLeuLysAsnCysSerValArgCysP 383

385 ..AAGGACACGGTTATGAGTATCTATTCGCCGACGGGAGGCGATCGTT 432
   :::::::::::::: ::::: :::::

```

```

383 heLysAspAsnGlyHisThrLeuLeuLeuGluAspSerIleSerLeuLeu 399
433 TTCACAGCTCATCTCAACTCGCGCATATAAGCCCAATTAAACCGCTCATCC 482
400 .....ThrValIleLysGlyThrGly..... 406
483 ACACCTTATAAAGCCCCAAAGGAAAGCCCTTTAGCTACATCAGC.... 528
   :: ::::::::::: :::::
407 .....LysTyrArgArgSerTrpArgTyrAspLeuValSerAsp 420
529 .....TATCTCCTTGGT.....CTC 543
420 heLeuProProSerLysGlyCluLeuAlaTyrAlaLeuAspGluValLeu 436
544 AGGAGCTTAGGAAGCGCATAAAGCTGTTTT.....GA 578
   :::::::::::::: :::::
437 GlyPheLeuArgAsnAlaValGlySerValPhePheSerThrMetGluAs 453
579 AGGTAAGGTAAAGCTAAAGGAGTCGAAGACATCGAAGCCGTACCC.... 624
   :::::::::::::: :::::
453 polyLysIle.....ValLysGlyLeuAlaGlyValProAspL 466
625 .....GTTTGGTGGCGCTGAACACGCGCTCTAATGCTCGGCATCGGAAG 669
   :: ::::::::::: :::::
466 ysGlyProValLeuLeuValGlyTyrHisMetLeuMetGlyLeuGluLeu 482
670 CTTCTCTTATGATCTCAAGAACTGCGGAGCTGGATAGAGGACACAGGA 719
   ::::::::::: :::::
483 GlyProMetSerGlu.....AlaPheIleLysGluLys.. 493
720 CAACATTCTTCTA..... 732
   ::::::::::::::
494 AsnIleLeuPheArgGlyMetAlaHisProValLeuTyrSerAspAsnA 510
733 .....TACGGCACCATATAGAGTTCATTCGCTAT 762
   :::::::::::
510 spProAlaLysAlaPheAspTyrGlyAspTrpIleLysValPheGly... 525
763 AGGACATTGCAGGCTACAGAAATGAGTTCGAGGATATTAGAGGTTAT 812
   :::::
526 .....AlaTyrProValThrAlaThrAsnLeuPheLysLeuLe 538
813 ACAGAGCTCAACTCGGAACGTGTCCTCCCTCA.....GAGCTGA 853
   :::::
538 uAspSerLysSerHisValLeuLeuPheProGlyGlyAlaArgGluAlaL 555
854 AGCAGGTGGAAGGAGCTCTACTTACGAGCTTCGAGTTGGCACCAGAT 903
   :::::
555 euHisAsnArgGlyGluGlnTyr..... 562
904 AAGAGCTTAGGATATGAGAGAGGACGAAAGGACGCAAGACTT 948
   :::::::::::
563 .....LysLeuIleTrpProGluGlnGlnGluPheValArgMet 575

seq_name: /STD1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:ABG18270
seq_documentation_block:
ID ABG18270 standard; Protein; 936 AA.
XX
AC ABG18270;
XX
XX 18-FEB-2002 (first entry)
DT
XX
DE Novel human diagnostic protein #18261.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.

```





OM of: US-09-886-400-3 to: SPTREMBL\_19:\* out\_format : pfs  
 Date: Jun 11, 2002 10:39 PM  
 About: Results were produced by the GenCore software, version 4.5.  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL=framet-n2p.model -US09886400/tunat\_11062002.150916.17416/app\_query.fasta\_1.1169  
 -O=/cgn2.1/USPTC\_SPOOL/US09886400 -DEV=xlh  
 -DB=SPTREMBL\_19 -QEMT=fastan -SUFFIX=rspt -GAPO=12.000  
 -GAPEXT=4.000 -MINWATCH=0.100 -LOOCL=0.000 -LOOPEXT=0.000  
 -QAPO=4.500 -GAPEXT=0.050 -XGAPO=10.000 -XGAPEXT=0.500  
 -FGAPO=6.000 -FGAPEXT=7.000 -YGAPO=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
 -TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcp  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM\_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09886400.cgn1.1.219 -NCPU=6 -ICPU=3 -LONGLOG  
 -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPXY -WAIT\_THREADS=1

## Search information block:

Query: US-09-886-400-3  
 Query length: 1095  
 Database: SPTREMBL\_19:\*  
 Database sequences: 562222  
 Database length: 172994329  
 Search time (sec): 61.230000

Sequence	Strid Orig	ZScore	EScore	Len	Documentation	
sp_archaea:Q9HHB5	+	1532.00	2383.59	56e-125	364	Q9HHB5 Pyrococcus furiosus, alp
sp_bacteriap:O58106	+	1494.00	2324.17	1.1e-121	364	Q58106 Pyrococcus horikoshii, h
sp_bacteriap:P74630	+	162.50	238.26	1.2e-05	529	P74630 Synechocystis sp. (strai
sp_bacteriap:Q97BM4	+	149.50	221.18	0.0001	378	Q97BM4 Thermoplasma volcanium,
sp_bacteriap:Q97GF3	+	147.00	214.06	0.0003	527	Q97GF3 Clostridium acetobutylic
sp_bacteriap:Q973T0	+	142.50	208.70	0.0006	443	Q973T0 Sulfolobus tokodaii, hyp
sp_bacteriap:Q972N0	+	136.50	192.52	0.0025	895	Q972N0 Sulfolobus tokodaii, hyp
sp_bacteriap:Q50094	+	130.50	187.62	0.0074	560	Q50094 Pyrococcus horikoshii, h
sp_bacteriap:Q9K004	+	128.00	178.97	0.0139	923	Q9K004 Bacillus halodurans, bhl
sp_fungi:Q9C105	+	127.50	175.32	0.0164	1236	Q9C105 Schizosaccharomyces pom
sp_bacteriap:Q9V294	+	124.00	168.91	0.0340	1362	Q9V294 Pyrococcus abyssi, anyl
sp_bacteriap:Q9H191	+	122.50	179.51	0.0332	357	Q9H191 Thermoplasma acidophilum
sp_bacteriap:Q30246	+	120.50	177.32	0.0485	324	Q30246 Archaeoglobus fulgidus,
sp_bacteriap:Q9V0M7	+	119.00	168.98	0.0760	602	Q9V0M7 Pyrococcus abyssi, hypot
sp_bacteriap:Q83377	+	111.00	157.78	0.3663	526	Q83377 Treponema pallidum, cons
sp_bacteriap:Q972D2	+	109.50	157.01	0.4759	447	Q972D2 Sulfolobus solfataricus,
sp_invertebrate:Q9BKV7	+	108.50	144.93	0.7555	1325	Q9BKV7 Leishmania major, pp9
sp_fungi:Q94317	+	108.00	152.94	0.6710	534	Q94317 Schizosaccharomyces pom
sp_fungi:Q96WV6	+	106.50	131.18	1.47	3971	Q96WV6 Schizosaccharomyces pom
sp_bacteriap:Q53278	+	106.00	149.96	0.9986	526	Q53278 Mycobacterium tuberculosis
sp_bacteriap:Q97Y0	+	105.50	143.96	1.26	902	Q97Y0 Sulfolobus solfataricus,
sp_plant:Q9VND2	+	104.00	144.01	1.60	704	Q9VND2 Arabidopsis thaliana (mc
sp_plant:Q9WBD0	+	103.50	147.39	1.59	458	Q9WBD0 Prunus pyrifolia (japanes
sp_bacteria:Q93MG7	+	103.50	143.27	1.77	701	Q93MG7 Thiobacillus ferrooxidan
sp_invertebrate:Q13131	+	103.00	141.78	1.99	754	Q13131 Oncoerythrus mykiss (rai
sp_invertebrate:Q13131	+	102.00	140.22	2.43	754	Q13131 Oncoerythrus mykiss (rai
sp_bacteriap:Q84233	+	101.00	144.43	2.57	415	Q84233 Chlamydia trachomatis, n
sp_bacteriap:Q97M42	+	100.50	147.96	2.55	266	Q97M42 Clostridium acetobutylic
sp_bacteriap:Q83182	+	100.50	139.78	3.13	619	Q83182 Treponema pallidum, alph
sp_fungi:Q74851	+	99.50	132.59	4.42	1131	Q74851 Schizosaccharomyces pom
sp_invertebrate:Q88738	+	99.00	117.53	6.94	4845	Q88738 Mus musculus (mouse), u
sp_invertebrate:Q17585	+	98.00	137.88	4.92	503	Q17585 Caenorhabditis elegans
sp_bacteriap:Q67347	+	98.00	133.94	5.44	764	Q67347 Aquifex aeolicus, hypoth
sp_invertebrate:Q9Y076	+	98.00	132.55	5.61	873	Q9Y076 Leishmania major, prot
sp_invertebrate:Q9Y075	+	97.50	139.74	5.09	383	Q9Y075 Leishmania major, prot
sp_bacteria:Q88779	+	97.50	139.27	5.15	402	Q88779 Yersinia pestis, transp
sp_bacteriap:Q92DR2	+	97.50	135.31	5.68	605	Q92DR2 Rickettsia prowazekii, r
sp_plant:Q9ZQF1	+	97.50	134.97	5.73	627	Q9ZQF1 Arabidopsis thaliana (mc
sp_plant:Q9M2S4	+	97.50	134.13	5.85	684	Q9M2S4 Arabidopsis thaliana (mc
sp_bacteriap:Q98PQ2	+	97.50	131.50	6.23	888	Q98PQ2 Mycoplasma pulmonis, iso

seq\_name: sp\_archaea:Q9HHB5  
 seq\_documentation:Q9HHB5  
 ID Q9HHB5 PRELIMINARY; PRT; 364 AA.  
 AC Q9HHB5;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE ALPHA-GALACTOSIDASE.  
 GN GALA.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=2261;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 3638;  
 RA Verhees C.H.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF195244; AAG28455.1;  
 DR InterPro: IPR004300; Glyco\_hydro\_57.  
 DR Pfam: PF03055; Glyco\_hydro\_57; 1  
 DR SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;  
 SQ SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;

alignment\_scores:  
 Quality: 1532.00 Length: 363  
 Ratio: 4.441 Gaps: 0  
 Percent Similarity: 95.041 Percent Identity: 79.063  
 alignment\_block:  
 US-09-886-400-3 x Q9HHB5  
 Align seg 1/1 to: Q9HHB5 from: 1 to: 364  
 1 TTGAGAGCGCTCGTCTTTCCAGGCAACCTCCAGTATCCGAAATCCCAAA 50  
 1 MCAAGAlaLeuValPheHisGlyAsnLeuGlnTrpAlaGluLeuProly 17  
 51 GAGCAATCCCAAGGTCATAGAGAGGATACATCCAGTATCCAGGATCAGA 100  
 17 SSerGluLeuProlysValIleGluLysAlaTrpPheProThrIleSerG 34  
 101 CACTGATTAAAGAGAAATTCCTTTGGGCTCAACATAACGGCTATACC 150  
 34 LuLeuIleArgArgGluLeuLeuProPheGlyLeuAsnIleThrGlyTrpSer 50  
 151 TTAAAGTTCTCCCGGAGGATATTATAGACCTCGTTAAAGGGGCGATCGC 200  
 51 LeuSerPheLeuProlysAspLeuIleAlaLeuIleLysGluGlyIleG 67  
 201 GAGTACCTGTAGAGATAATTCGAGACAGTACACGACCAATATCTCC 250  
 67 uSerGlyLeuIleGluLeuGlyThrSerTrpThrHisAlaIleLeuP 84  
 251 CCCTCTCTCCGCTTACGAGATAGAGACCAAGTTCAGAGATAGGGAAC 300  
 84 roLeuLeuProLeuSerArgValGluAlaGlnIleLysArgAspArgGlu 100  
 301 GTTAAGGAGAGCTCTTCGAGCTTTCTCCAAAGGATTCCTCGCTGCCAGA 350  
 101 ValLysGluAsnIleLeuGluValSerProGluGlyPheTrpLeuProG 117  
 351 GCTCGCTATGACCGAATATCCCTGCCATATCTAGAGCAACAGGTTATG 400  
 117 uLeuAlaTrpAspProIleIleProAlaIleLeuArgAspAsnAsnTrp 134  
 401 AGTATCTATTCGCCGACGGGGGCGGATGCTTTTCTCAGCTCATCTCAAC 450

134 lutyLeuPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150  
451 TCSCGCGATAAGCAATTAACCGCTCTTCCACACCTTATTAAGGCCCA 500  
151 SerAlaileysProleuysProleuysProleuysProleuysProleuys 167  
501 AAGGAAAACGGTTTAGGTACATCAGCTATCTCCCTGCTCAGGGAGC 550  
167 nArgGlyGluGlyLeuValtyrLeuAsnTyrrLeuLeuGlyLeuArgGlu 184  
551 TTAGGAGGCGATAAAGCTCGTTTGAAGCTAAGCTAAGCTAAGGCA 600  
184 eulysLysAlaileAsnLeuValPheGluGlyLysValThrLeuGluAla 200  
601 GTCAAAGACATCGAAGCGCTACCGCTTGGTGGCGTGAACACGGCTGT 650  
201 VallysGluileGluAlaileProValTrpValSerIleAsnThrAlaVa 217  
651 AATGCTCGGATCGGAAGGCTTCTCTTATGATCCTAAGAAAGTGGCA 700  
217 lMetLeuGlyAlaGlyArgPheProleuMetAsnProLysLysValAla 234  
701 GCTGATAGAGCAAGCAACATCTTCTATACGGCACCGCATATAGAG 750  
234 ystrPvalLysGluLysAspGluileLeuLeuTyrrGlyThrAspIleGlu 250  
751 TTCATTGGCTATAGGACATTCAGGCTACAGNATGAGTGTGAGGGATT 800  
251 PheLeuGlytyrArgAspIleAlaGlytyrLysIleThrIleSerAsnLe 267  
801 ATTAGAGGTTATAGCAGCTCAACTCGGAACCTGTCCCTCCCTCAGAGC 850  
267 uLeuGluileleAsnGluLeuGluGlyGluLeuGlyLeuProArgLysI 284  
851 TGAAGCAGTGAAGGAGGACTCTACTACGGACTTCGAGTTCGAGTGGCA 900  
284 lelysHisSerGluLysLysLeuTyrrLeuArgThrSerTrpAlaPro 300  
901 GATAAGAGCTGAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 950  
301 AsplysSerLeuArgIleTrpThrGluAspGluGlyAsnAlaArgLeuAs 317  
951 TATGCTGCTACATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1000  
317 nMetLeuThrSerTyrrMetAspGlyGluLeuAlaPheLeuAlaGluAsn 334  
1001 CGGATGCAAGGGGATGGAGCCCTCCCTCAGAGGAGGAGGAGGAGGAG 1050  
334 erAspAlaArgGlyTrpGluProleuProGluArgLeuAspAlaPhe 350  
1051 CGGCGGATATATACGATTGGAGGGGAGGAGGAGGAGGAGGAGGAG 1089  
351 LysAlailetyrThrHisTrpArgserGluAsnGlyLys 363  
seq\_name: sp\_archaeap:058106

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
RA Masuchi Y., Shizuya H., Kikuchi H.,  
RT "Complete sequence and gene organization of the genome of a hyper-  
RL thermophilic archaeobacterium, Pyrococcus horikoshii OT3.",  
RL DNA Res. 5:55-76(1998).  
DR EMBL: AP000002; BAA29442.1;  
DR InterPro: IPR004300; Glyco\_hydro\_57\_1;  
DR Pfam: PF03065; Glyco\_hydro\_57\_1;  
KW Hypothetical protein; Complete proteome  
SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD CRC64;  
alignment\_scores:  
Quality: 1494.00 Length: 363  
Ratio: 4.381 Gaps: 0  
Percent Similarity: 93.939 Percent Identity: 75.758  
alignment\_block:  
US-09-886-400-3 x 058106  
Align seg 1/1 to: 058106 from: 1 to: 364  
1 TTGAGAGCGCTCGCTCTTTCACGGCAACCTCCAGTATCCGAAATCCCAA 50  
1 MetArgAlaLeuilePheHisGlyAsnLeuGlnTyrrAlaGluileProly 17  
51 GAGCGAAATCCCAAGGTGCATAGAGAGGATACATCCAGTCATCGAGA 100  
17 sHisGluileSerLysValileGluLysSerTyrrPheProThrIleSerg 34  
101 CACTGATTAAAGAGAAATTCCTTTGGCTCAACATACGGGCTATACC 150  
34 lueuileLysArgGluileProPheGlyLeuAsnIleThrGlyTyrrSer 50  
151 TTAAGTCTCCCGAGGATATATAGACCTCGTTTAAAGGGGCGATCCG 200  
51 LeuGlnPheLeuProGlnGluLeuileHisLeuileLysGluGlyileG 67  
201 GAGTGACCTCATAGATATATCGGAACGAGCTACACGACGCAATACCTC 250  
67 uSerGluileleGluileuGlyThrSerTyrrHisAlaileLeuP 84  
251 CCTCTCCCGCTTAGCAGAGTAGAAGCAAGTTCAGAGAGATAGGAA 300  
84 roLeuLeuThrLeuSerArgileGluAlaGlnileLysArgAspArgGlu 100  
301 GTTAAGAGAGAGCTCTCGAGCTTCTCCAAGGGATTCGGCTCCAGA 350  
101 ileLysGluilePheGluValSerProGlyGlyPheTrpLeuProG 117  
351 GCTCGCTATGACCCGATATCCCTGCCATCTGAAAGGACACGTTATG 400  
117 uLeuAlaTyrrAspProIleleProAlaileLeuArgAspAsnGluTyrg 134  
401 AGTATCTATTCCGCGGAGGCGGATGCTTTCTCAGCTCATCTCAAC 450  
134 lutyLeuPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150  
451 TCGCGGATAAAGCAATTAACCGCTCTTCCACACCTTATTAAGGCCCA 500  
151 SerAlaileysSerileLysProLeuTyrrProTyrrLeuileLysAlaG 167  
501 AAGGAAAACGGTTTAGGTACATCAGCTATCTCCCTGCTCAGGGAGC 550  
167 nArgGlyGluGlyPheValtyrLeuAsnTyrrLeuLeuGlyLeuArgGlu 184  
551 TTAGGAGGCGATAAAGCTCGTTTGAAGCTAAGCTAAGCTAAGGCA 600  
184 eulysLysAlaileAsnLeuThrPheGlyGlyLysValThrLeuGluAla 200

seq\_documentation\_block:  
ID 058106 PRELIMINARY; PRT; 364 AA.  
AC 058106;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHEICAL 41.8 KDA PROTEIN PH0368.  
GN PH0368.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,





```

731 TATAGGACCGATATAGAGTTTCATGGCTATAGGACATTCGAGGCTAC 780
  :: ||||| ::||| ::||| ::||| ::|||
227 alasn.....LeuPheMetAspTyrGluThrPheGlyGluHis 239
  :: ||||| ::||| ::||| ::||| ::|||
761 AGAATGAGTGTGAGGATATAGAGTTTATAGACGAGCTC..... 822
  :: ||||| ::||| ::||| ::||| ::|||
240 GlnThrGlnGluThrGlyIlePheAspPheMetLysTyrLeuProValty 256
  :: ||||| ::||| ::||| ::||| ::|||
823 .....AACTCGGAACGTGCCTTCCTCCCTCAGAGCTGA 853
  ::||| ::||| ::||| ::||| ::|||
256 rPheArgAspTyrGlyIleGluThrIleThrIleSerGluAlaGluLysA 273
  ::||| ::||| ::||| ::||| ::|||
854 AGCACAGTGGAGGAGCTCTACTTA.....CGGACTTCGAGTTGGGCA 897
  ::||| ::||| ::||| ::||| ::|||
273 rGHisArgValLysAspValLeuSerIleProGluThrIleSerIlePala 289
  ::||| ::||| ::||| ::||| ::|||
898 CCAGAT...AAGAGCTTGAGGATATGAGAGGAGGAGAA 933
  ::||| ::||| ::||| ::||| ::|||
290 AspThrArgArgAspLeuSerAlaTrpLeuGluAsnGlu 302
  ::||| ::||| ::||| ::||| ::|||
seq_name: sp_bacteriap:Q97GF3
seq_documentation_block:
ID Q97GF3 PRELIMINARY; PRT; 527 AA.
AC Q97GF3;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE UNCHARACTERIZED CONSERVED PROTEIN, RELATED TO
DE ALPHA-AMYLASE/ALPHA-MANNOSIDASE.
GN CAC2414.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007742; AAK80369.1;
DR Inter-Pro: IPR003803; DUF200.
DR Pfam: PF02651; DUF200; 1.
KW Complete proteome.
SQ SEQUENCE 527 AA; 61485 MW; 03706E16907929C6 CRC64;

```

```

alignment_scores:
  Quality: 147.00 Length: 486
  Ratio: 0.766 Gaps: 20
  Percent Similarity: 39.506 Percent Identity: 18.930
alignment_block:
US-09-886-400-3 x Q97GF3 ..
Align seg 1/1 to: Q97GF3 from: 1 to: 527
13 GTCITTCAGCGCACTCCAGTATGCGGAATCCCAAGAGCGAA..... 57
  ::||| ::||| ::||| ::||| ::|||
9 ValLeuHisSerHisMetProPheValArgHisProGluThrArgAspse 25
  ::||| ::||| ::||| ::||| ::|||
58 .....ATCCCAAAGGTCATAGAGAGGCATACATCCAG 91
  ::||| ::||| ::||| ::||| ::|||
25 rleuGluGluArgTrpLeuPheGluAlaMetSerGluCysTyrIleProL 42
  ::||| ::||| ::||| ::||| ::|||
92 TCATCGAG.....ACATGTATTAAAGAAATTCCTTTTGGG... 129

```

```

::||| ::||| ::||| ::||| ::|||
42 euileGluValTyrAspAsnLeuLeuLysAspAsnIleLysPheArgMet 58
  ::||| ::||| ::||| ::||| ::|||
129 .....CTCAACATAGCGGCTATACCTTAAGAGTTCCTCCCGA 166
  ::||| ::||| ::||| ::||| ::|||
59 ThrMetSerIleThrProLeuMetSerMetLeuGlnAspGluTyrLe 75
  ::||| ::||| ::||| ::||| ::|||
130 .....CTCAACATAGCGGCTATACCTTAAGAGTTCCTCCCGA 166
  ::||| ::||| ::||| ::||| ::|||
75 uasSerArgTyrLeuAsnTyrLeuLysTyrThrIleGluLeuSerGluL 92
  ::||| ::||| ::||| ::||| ::|||
167 AGGATATTATA..... 177
  ::||| ::||| ::||| ::||| ::|||
92 ySGluIleLeuArgThrLysAsnAsnArgGluGluAsnLysValAlaLeu 108
  ::||| ::||| ::||| ::||| ::|||
177 ..... 177
  ::||| ::||| ::||| ::||| ::|||
109 PheTyrAsnLysArgAlaGluAsnThrLeuLysIleTyrGluLysTyrAs 125
  ::||| ::||| ::||| ::||| ::|||
178 ....GACCTCGTTAAAGGGGCATCGCGAGTGAC.....CTGATAG 214
  ::||| ::||| ::||| ::||| ::|||
125 pasAsnLeuIleAsnAlaPheArgLysTyrAspArgLeuGlyCysValG 142
  ::||| ::||| ::||| ::||| ::|||
215 AGATAATCGGAACGAGCTACACGACCAATACCTCCCTCTCCCGCTT 264
  ::||| ::||| ::||| ::||| ::|||
142 luileIleThrCysAlaAlaThrHisAlaLeuProLeuIleLeuIle 158
  ::||| ::||| ::||| ::||| ::|||
265 AGCAGA.....GTGAAGCAGCAGTTCCAGAGATAGGAGCTTAAGGA 308
  ::||| ::||| ::||| ::||| ::|||
159 AsnArgGlnAlaValLysAlaGlnIleAlaThrGlyValGlnSerTyrI 175
  ::||| ::||| ::||| ::||| ::|||
309 AGAGCTCTTCGAGCTTTCCTCAAAGGATTCCTGCTGCCAGAGCTCGCT 358
  ::||| ::||| ::||| ::||| ::|||
175 eAsnThrMetGlyHisGluProAsnGlyIleTrpLeuProGluCysAla 192
  ::||| ::||| ::||| ::||| ::|||
359 ATGACCCGATATCCCTGCCATACCTAGCAGCACACGGTTATGATCTA 408
  ::||| ::||| ::||| ::||| ::|||
192 yrThrTyrGlyIleAspAsnIleLeuSerGluPheGlyIleLysTyrPhe 208
  ::||| ::||| ::||| ::||| ::|||
409 TTCGCCAGCGGAGGAGGATGCTTTCTCAGCTCATCTCAACTCGCGCAT 458
  ::||| ::||| ::||| ::||| ::|||
209 IleSerGluGlyLysAlaIleAspTyrAlaSerPro..... 220
  ::||| ::||| ::||| ::||| ::|||
459 AAGCCCAATTAACCGCTCTATCCACACCTTATAAGGCCCAAGGAAA 508
  ::||| ::||| ::||| ::||| ::|||
221 .....LysProMetTyr..... 224
  ::||| ::||| ::||| ::||| ::|||
509 AGCGCTTTAGGTACATCAGCTATCTCCTTGGTCTCAGGGAGCTTAGGA 558
  ::||| ::||| ::||| ::||| ::|||
224 ..... 224
  ::||| ::||| ::||| ::||| ::|||
559 GCGATAAAGCTCGTTTTTGAAGGTAAGGTAACGCTAAAGCGACGTCAA 608
  ::||| ::||| ::||| ::||| ::|||
224 ..... 224
  ::||| ::||| ::||| ::||| ::|||
609 CATCGAAGCCGTACCGCTTTGGTGGCGGTGAACACGCGTGTATGCTCG 658
  ::||| ::||| ::||| ::||| ::|||
225 .....GlyThrAsnThr..... 228
  ::||| ::||| ::||| ::||| ::|||
659 GCATCGGAAGGCTTCCTCTTATGATCTTAAGAAAGTGGCGAGCTGGATA 708
  ::||| ::||| ::||| ::||| ::|||
229 .....ProIleAlaAlaProSerGlyValCysAlaPheGly 240
  ::||| ::||| ::||| ::||| ::|||
709 GAGGACAGGACAAACATTTCTATACGGCAGCATATAGATTTCATTGG 758
  ::||| ::||| ::||| ::||| ::|||
241 ArgaspMetaspSer.....SerTyrGlnValTrpSeraspPheMetGI 255
  ::||| ::||| ::||| ::||| ::|||
759 C.....TATAGGACATTGCAGGCT 778
  ::||| ::||| ::||| ::||| ::|||
255 yTyrProGlyAspPheAsnTyrArgGluPheTyrArgAspile...GlyP 271
  ::||| ::||| ::||| ::||| ::|||
779 ACAGATGAGTGTTCAGGAGTATTATAGAGGTTATAGACGAGCTCAACTCG 828
  ::||| ::||| ::||| ::||| ::|||

```

```

271 hecluleuPrometGlutrylIleLysProTyrIleAsnGlu.....Asn 285
829 GRACGTGCCCTCCCTCAGAGCTGAGCAC.....AG 860
286 GlyIleArgIleAspThrGlyPheLysTyrTyrLysIleThrGlyAsnSe 302
861 TGAAGGGAG...CCTACTACGGACTCGAGT.....TGGG 895
302 rGlyGluLysGlyIleTyrAsnArgGluAsnAlaMetLysLysValIrpG 319
896 CACCAAGTAAGAGCTTGGAGTATGCGAGAGAGGAGGAGGAAC...GCA 942
319 luhisAlaSerHisPheAlaSerCysArgHisAspGlnIleAsnAlaAla 335
943 AGACTTAATATG..... 954
336 AlaAlaAsnMetAspLysProIleIleThrCysProTyrAspThrG1 352
954 ..... 954
352 uLeuTyrGlyHisTyrTrpPheGluGlyProAspPheIleAsnAlaPheI 369
955 .....CCTGCTCAATATGAGGGGCGAATC 981
369 leArgLysSerAlaGluAspTrpThrSerTyrGluLeuIleThrProThr 385
982 GCCCTTTTACCGAGAACAGCATGCAAGGGATGGAGGCCCTCCCTGA 1031
386 GluTyrLeuLysAsnAsnSerMetValGlnCysSerSerProSerPro.. 401
1032 GAGGAGCGTGCATGCCCTCCGGCGATATATACGATTGGAGGGGTGAAA 1081
402 .....SerSerTrp...GlyGluA 407

1082 ATGGGGAA 1089
407 snGlyAsp 409

seq_name: sp_archaeap:Q973T0
seq_documentation_block:
ID Q973T0 PRELIMINARY; PRT; 443 AA.
AC Q973T0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEICAL PROTEIN ST0817.
GN ST0817.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yanagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7."
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000983; BAB65830.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 443 AA; 53218 MW; 497087264BC03C4 CRC64;

```

```

alignment_scores:
Quality: 142.50 Length: 343
Ratio: 0.902 Gaps: 15
Percent Similarity: 46.064 Percent Identity: 23.032

```

```

alignment_block:
US-09-886-400-3 x Q973T0
Align seg 1/1 to: Q973T0 from: 1 to: 443

64 AAGGTCAATAGAGAGGATACATCCAGTATCGAGACACTGATTAAGA 113
61 AAGVallysAsnLysCysTyrIleProAlaThrArgIleLeuLeu...G1
114 AGAATATCTCTTTGGCTCAACATAACGGGCTATACCTTAAGTCTC. 162
61 ucluliegluasngly...GluserGluGlyTyrAspPheLysPhePhep 77
163 .....CCGAAGGAT 171
77 heSerLeuSerGlyThrPheIleGluGlnAlaGluLysTrpGlyLysAsp 93
172 ATATAGACCTCGTTAAAGGGGGCATCCGAGTGNACCTGATAGAGATAAT 221
94 ValLeuGluLeuPheGlnGlnLeuSerTyrThrHisLysValGluPheLe 110
222 CGGAAGGAGCTACACGACACCAATACTCCCTCCTC...CCGTTAGCA 268
110 uSerGlnThrTyrTyrHisSerIleThrSerLeuTrpGluAspLeuThrG 127
269 GAGTAGAGACCAAGTTTACAGAGATAGGAGTAAAGAGAGCTCTTC 318
127 lutrPrpLysGluGlnValGlnMetHisLysAsnLeuIleArgAspTyrPhe 143
319 GAGCTTTCCTCAAGGAGTCTGGCTGCGCAGAGCTGCGCTATGACCGGAT 368
144 GlyGlnGluProValThrPheGluAsnThrGluLeuLeuLeuThrProAr 160
369 AATCCCTGCCATCTAGAGGACACAGGTTATGAGTATCTATTCGCCGACG 418
160 grileValLysGluIleGluLysLeuGlyPheLysValValIleThrGluG 177
419 GGGAGGCGATGCTTTTCTCAGCTCATCTCACTCGGCGATAAAGCCAATT 468
177 lLys.....GluserLeuLeuLysGlyLys 185
469 AAACCGCTCTATCCACACCTTATAAGGCCCAAGGAAAGCGCTTTAG 518
186 SerPro.....AsnArgValTyrArgIleAr 194
519 GTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAGGCGATAAAGC 568
194 gaspThrLysLeuSerLeuLeuLeuArgAsnTyrArgLeuSerAspAspi 211
569 TCGTTTGTGAAGTAAGTAAAGCTAAAGGCGAGTCAAAAGACATCGAAGCC 618
211 leAlaPheArg.....PheSer 216
619 GTACCGCTTTGGTGCCCTGACACGCGCTGTAACTCGCGCATCGGAAG 668
217 AsnProLysTrp.....AspG1 222
669 GCCTCCTCTATGAATCCCTAAGAAAGTGGCGAGCTGGATAGAGGACAAG 718
222 nTyrProLeu...ThrAlaSerLysPheAlaAspTrpValSerTrpSerG 238
719 ACAACATTTCTTATACGCCACCGCATATAGATTCATTGGCTATAGGAC 768
238 lGlyGluIle.....GlyLeuIlePheValAspTyrGluThr 250
769 ATGTCAGGCTACAGATAGTGTGTAGGGGATTTATAGAGTTATAGACGA 818
251 PheGlyCluHisThrProGluSerGlyIleLeuAspPheLeu..... 265
819 GCTCAACTCGGAAGTGTGCTTCCCTCAGAGCTGAGCAGCTGGA.... 864
266 .....ArgTrpLeuProArgGluLeuHisArgArgGlyValG 278

```

```

865 .....AGGAGCTCTACTTAGG..... 882
|||||:|||||
278 luPheLysLeuProArgGluValTyrLysGlnTyrTyrAspGluIleVal 294
|||||:|||||
883 .....ACTTCGAGTTGGCA..... 909
|||||:|||||
295 IleAspSerThrValSerTrpAlaAspIleAsnLysAspGluSerTr 311
|||||:|||||
910 .....TTGAGGATATGGAGAGGACGAGGAAAGCAAGACTTA 949
|||||:|||||
311 pLeuGlyAsnIleMetGlnTrpAlaTyrAspGluMetValArgArgThrG 328
|||||:|||||
950 ATATGCTGCTCCTACATATAGGGGGGAA 978
|||||:|||||
328 luMetLeuAlaLysGluAlaGlyGlyGlu 337
|||||:|||||
seq_name: sp_archaeap:Q972N0
seq_documentation_block:
ID Q972N0 PRELIMINARY; PRT; 895 AA.
AC Q972N0;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST1102.
GN ST1102.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7."
RL DNA Res 8:123-140(2001).
DR EMBL; AP000985; BAB56135.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 895 AA; 101064 MW; 6BC7CD8380DFBEC CRC64;

```

```

alignment_scores:
  Quality: 136.50      Length: 410
  Ratio: 0.669        Gaps: 20
  Percent Similarity: 49.756      Percent Identity: 20.732

alignment_block:
US-09-886-400-3 x Q972N0 ..

Align seg 1/1 to: Q972N0 from: 1 to: 895

7 GCGCTGCTTTTCAGCGCACTCCAGTATCGCGAAATCCCAAGAGCGGA 56
|||||:|||||
377 SerLeuValGlyAlaTyrGluLeuGlnAlaLeuLeuIleLysGlnPheAs 393
|||||:|||||
57 AATCCCAAGGTCATAGAGAGGACATACATCCAGTCATC..... 96
|||||:|||||
393 nValSerValThrIleAsp.....PheThrProValLeuLeuTyrGlnT 408
|||||:|||||
97 ..GAGACACTGATTAAGAAGAA.....ATTCCTTTTGGG 129
|||||:|||||
408 rpGluThrIleLeuHisGluGluAsnAlaThrPheThrSerAsnPheGly 424
|||||:|||||
130 CTCACATAACG.....GGCTATACCTTAAGATTCCT 161
|||||:|||||
425 ValAsnValSerHisAspIleGluAlaValAsnTyrThrLeu..... 438
|||||:|||||

```

```

162 CCCGAAGATATTATAGACCTCGTTAAAGGGGCATCGCGAGTGAACCTGA 211
|||||:|||||
439 .....AsnLeuTyrLysThrLeuIleLysGluGlyLysV 450
|||||:|||||
212 TAGAGATAATCGGAAGAGGTACACGACGCAATACCTCCCTCCCTC... 258
|||||:|||||
450 alGluValLeuThrValProPheTyrHisProLeuGlnProLeuLeuLeu 466
|||||:|||||
259 .....CCGCTTAGCAGAGTAGAAGACACAAAGTTTCAGAGATAGGGA 299
|||||:|||||
467 GlnAspGlyTyrTrpSerAspValLeuAlaGlnIleGluMetGlyGluAs 483
|||||:|||||
300 AGTTAAGGAAGAGCTCTTCGAGCTTTTCCAAAGGGATTCTGGCTGCCAG 349
|||||:|||||
483 nMetThrHisGluValPheGlyValTrpAlaAsnGlyThrTrpThrProG 500
|||||:|||||
350 AGCTCGCTATGACCCGATAATCCCTGCCATCTGAAGAGGACACACGGTTAT 399
|||||:|||||
500 luMetAlaPheAspMetAspLeuValGlyLeuTyrAsnGluSerGlyIle 516
|||||:|||||
400 GAGTATCTATTCGCCGACGGGGGCGATGCTT...TTCTCAGCTCATCT 446
|||||:|||||
517 SerPheThrIleLeuAspGlnGlnAlaPheLeuProTyrValThrLeuVa 533
|||||:|||||
447 CAACTCGCGGATAAAGCCAAATTAACCGCTCTATCCACACCTTATAAGG 496
|||||:|||||
533 lAsnGlySerLeuAsnProAspGlnProPheIleValGlu..... 546
|||||:|||||
497 CCCAAAGGAAAGCGCTTTAGGTACATCAGCTATCTCTTGCTCTCAGG 546
|||||:|||||
547 .....Asn 547
|||||:|||||
547 GAGCTTAGGAAGCGGATAAAGCTCGTTTGAAGTAAGTAAGCTAA 596
|||||:|||||
548 AsnLeuGlyGlnThrIleIleValLeuPheArg...AsnThrThrLeuSe 563
|||||:|||||
597 G.....CCATCAAGACATCGAAGCCGCTACCCGTTGGTGCGCG 637
|||||:|||||
563 rAsnGluPheGlyPheLysPheSerGlnSerProGlnLeuThrAla.. 579
|||||:|||||
638 TGAACACGCTGTAATGCTCGGCATCGGAAGGCTCTCTTATGAATCCT 687
|||||:|||||
580 .....GlnGluLeuIleGlnGlnLeuAlaGluIleTyrMetAsnAsnPro 594
|||||:|||||
688 AAGAAAGTGGCGAGCTGGATAGAGGACAGACACATCTTCTTATACGG 737
|||||:|||||
595 GlyGlyValValThrValAlaLeuAspGlyGluAsnProLeuIlePheAs 611
|||||:|||||
738 C.....ACCGATATAGAGTTTCATGCTATAGGACATG 772
|||||:|||||
611 nProThrThrGlyProSerAspLeuTyrAlaIle...TyrGlnAlaLeuS 627
|||||:|||||
773 CAGGCTACAGAAATAGTGTGTAGGGATATTAGAGTTATAGACGAGCTC 822
|||||:|||||
627 erGluTyrGlnGlySer.....TrpLeuValThrGlnThr 638
|||||:|||||
823 RACTCGGAAGTGGCTTCCTCCCTCAGAGCTGAAGCACAGTGGGAAGGAGCT 872
|||||:|||||
639 AlaSerGluAlaIle.....AlaThrHisLysProThrSerIleIleTh 653
|||||:|||||
873 CTACTTACGAGCTTCGAGTTGGGCCAGATAGAGCTTGAGGATATGGA 922
|||||:|||||
653 rAsnLeuProValAsnSerTrp.....AspLeuAsnLeuAsnTyrTrp 668
|||||:|||||
923 GA.....GAGGACGAAGGGAAGCGCA 942
|||||:|||||
668 sAsnGlyTyrIleGlyLysThrGluIleTrpGlnAsnValSerLeuAla 684
|||||:|||||
943 AGACTTAATATGCTGCTCCTACAAATATAGGGGCGAAGCTCGCCCTTTTACG 992
|||||:|||||
685 ArgGluTyrLeuIleAlaTyrThrVal.....AlaLeuGlu 696
|||||:|||||

```



```

seq_name: sp_bacteriap:O9KD04

seq_documentation_block:
ID      O9KD04      PRELIMINARY;      PRT;      923 AA.
AC      O9KD04;
DT      01-OCT-2000 (TReMBLrel. 15, Created)
DT      01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT      01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE      BH1415 PROTEIN.
OS      Bacillus halodurans.
OC      Bacillus/Firmicutes; Bacillus/Clostridium group;
OC      Bacillus/Staphylococcus group; Bacillus.
NCBI_TaxID=86665;
[1]
RN
RP      SEQUENCE FROM N.A.
RC      STRAIN=C-125 / JCM 9153;
RX      MEDLINE=20512582; PubMed=11058132;
RA      Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Ma
RA      Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
RA      Horikoshi K.;
RA      "Complete genome sequence of the alkaliphilic bacterium Ba
RT      halodurans and genomic sequence comparison with Bacillus s
RL      Nucleic Acids Res. 28:4317-4331(2000).
DR      EMBL: AP001512; BAB05134.1; -.
DR      InterPro: IPR003803; DUF200.
DR      InterPro: IPR001296; Glycos_transf_1.
DR      Pfam: PF02851; DUF200; 1.
DR      Pfam: PF00534; Glycos_transf_1; 1.
DR      complete proteome.
SQ      SEQUENCE 923 AA; 106638 MW; 8C7B7D64F3603553 CRC64;
KW

```

## alignment\_scores:

Quality: 128.00 Length: 482  
 Ratio: 0.612 Gaps: 24  
 Percent Similarity: 43.361 Percent Identity: 20.124

## alignment\_block:

US-09-886-400-3 x Q9KD04 ..

Align seg 1/1 to: Q9KD04 from: 1 to: 923

```

7 GGGCTGCTTTCAGCGCAACTCCAGTATGCCGAATCCCAAGACGGA 56
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
7 SerLeuValLeuHisAlaHisLeuProTyrValArgHisGlnGluAs 23
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
57 A.....ATCCCAAGGTCATAGAGAAGGCATACATCC 88
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
23 parGLeuGluArgTrpLeuPheGluAlaMetSerGluThrTyrIleP 40
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
89 CAGTCATCGACAGACTGATTAAAGAAATTCCTTTTGGCTCAACATA 138
  | | | | | | | | | | | | | | | | | | | | | | | | | |
40 roLeuLeuTrpAlaLeuGluLysLeuProValLysHisAlaValThrIle 56
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
139 .....AC 140
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
57 SerPheThrProProValMetGluMetLeuSerAspProLeuValGlnTh 73
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
141 GGGCTAT.....ACCTTAAAGTTCCTCCCGAAGGAT. 171
  | | | | | | | | | | | | | | | | | | | | | | | | | |
73 rArgTyrLeuAsnHisLeuGluAsnThrGluGlnLeuLysLysGluG 90
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
172 .....ATTATAGACTCGTTAAA 189
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
90 LuLysArgThrAsnAspGlnArgThrGlnAsnLeuValGlnPheTyrLys 106
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
190 GGG..... 192
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
107 GlnArgTyrGluLysLeuLysAlaThrPheLeuGlnTrpAspArgAsnLe 123
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
193 .....GGCATCGGAGTGACCTGATAGAG.....ATAA 220
  | | | | | | | | | | | | | | | | | | | | | | | | | |
123 uLeuLeGlyPheArgSer.....LeuMetGluAsnGluGlnCysThrLeuM 139
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
221 TGGGAACGAGCTACAGCAGCAATACCTCCCTCCCTCCCTTAGCAGA 270
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
139 etThrSerAlaAlaThrHisAlaPhePheProTyrLeuLysThrLysGlu 155
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
271 ...GTAGAAGCAAGTTCAGAGATAGGGAAGTTAAGGAAGCTCTT 317
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
156 AlaIleArgAlaGlnValArgHisGlyIleAlaCysPheGluGlnHisPh 172
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
318 CGAGCTTTCCTCAAGGATTCGTGCTGCCAGAGCTCGCTATGACCCGA 367
  | | | | | | | | | | | | | | | | | | | | | | | | | |
172 eGlyLysLysProLeuGlyPheTrpLeuProGluCysAlaPheSerProG 189
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
368 TAATCCCTGCCATCTAAGGACACAGGTTATGATATCTATTCGCCGAC 417
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
189 lyValAspArgIleLeuPheGluGluGlyIleArgTyrThrPheValAsp 205
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
418 GGGGAGGCGATCTTTC.....TCAGTCATCTCAACTCGGCGAT 458
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
206 GluHisAlaValLeuThrAlaAspProThrProHisLysGlySerSer.. 221
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
459 AAAGCCAATTAACCGCTCTAT...CCACAC.....CTTATAAGG 496
  | | | | | | | | | | | | | | | | | | | | | | | | | |
222 .....AlaProIleTyrSerProHisGlyIleAlaLeuPheProA 235
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
497 CCCAAAGGAAAAGCGCTTAGTACATCAGCTATCTCTCGTCTCAGG 546
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
235 rgHisThrGluLeuSerAlaLysValTrpSerSerThrLeuGly..... 249
  ::::::::::::::::::::| | | | | | | | | | | | | | | |
547 GAGCTTAGGAAGCGGATAAAGCTCGTTTTTGAAGGTAAGGTAACGCTAAA 596
  ::::::::::::::::::::| | | | | | | | | | | | | | | |

```

```

250 .....:::| | | | | | | | | | | | | | | |
  :::| | | | | | | | | | | | | | | |
597 GGCA...GTCAAAGACATCGAAGCGGTACCGTTTG.....630
  :::| | | | | | | | | | | | | | | |
257 gcluphetyrArgAspIleAlaTyrAspArgGluTrpAspTyrIleLysP 274
  :::| | | | | | | | | | | | | | | |
631 .....GTGCCCTGAACACGCGCTGTAATGCTCGGC 660
  :::| | | | | | | | | | | | | | | |
274 roHisValHisLysAspGlyIleArgIleAspThrGly.....LeuLys 288
  :::| | | | | | | | | | | | | | | |
661 ATCGGAAGCTTCCTCTATGAATCCCTAAGAAA.....GTGGCGAG 701
  | | | | | | | | | | | | | | | | | | | | | |
289 TyrHisArgIleThrGlyHisThrGluGluLysAspLeuTyrValargGI 305
  :::| | | | | | | | | | | | | | | |
702 CTGGATAGAGGACAAG.....GACAAACATT 726
  :::| | | | | | | | | | | | | | | |
305 utrpAlaGluLysArgValGlnGluHisAlaAsnHisPheIleGlyAlaI 322
  :::| | | | | | | | | | | | | | | |
718 ..... 726
  :::| | | | | | | | | | | | | | | |
322 leHisHisGluIleAspGlnHisGlyGlnAsnPheProTyrVal 338
  :::| | | | | | | | | | | | | | | |
727 CTTCTATACGGCACCAGATATAGAGTTTCATTTGGCTATAGGACATTCAGG 776
  :::| | | | | | | | | | | | | | | |
339 MetValThrProPheAspAlaLeuPheGlyHisTrpPheGluGI 355
  :::| | | | | | | | | | | | | | | |
777 CTACAGAATGAGTGTGAGGATTATTAGAGTT.....ATAG 814
  | | | | | | | | | | | | | | | | | | | | | |
355 yProGluTrp...IleGluAlaLeuTyrGluGlnGlyAlaAspArgValS 371
  :::| | | | | | | | | | | | | | | |
815 AGGAGCTCAACTCGGAACTGTCCCTCCAGAGCTGAAGCACACAGTGA 864
  :::| | | | | | | | | | | | | | | |
371 erPheIleThrProGluLeuTyrLeuGlnArgHisTyrGln...AspPhe 386
  :::| | | | | | | | | | | | | | | |
865 AGGAGCTCTACTTACGAGCTTCGAGTTGGCCACCAGATAAGAGCTTGAG 914
  :::| | | | | | | | | | | | | | | |
387 GlnThrAlaHisValSerPheSerThrTrpGlyArgAspGlyTyrGlyHI 403
  :::| | | | | | | | | | | | | | | |
915 GATATGGAGAGAGGAGGAGGAGGAGCTTAATATGCTGCTCTACATA 964
  :::| | | | | | | | | | | | | | | |
403 svalTrpLeuAsnAspHis...AsnAlaTrpMetTyrArgHisTyrHisA 419
  :::| | | | | | | | | | | | | | | |
965 ATATGAGGGGGAAGTCCGCTTTTA..... 990
  :::| | | | | | | | | | | | | | | |
419 rgMetGluLysAspLeuAlaLysIleValAlaMetTyrProGlnProThr 435
  :::| | | | | | | | | | | | | | | |
991 .....GCCGAGAACAGCGATGCAAGGGATGGGAGCCCT 1025
  :::| | | | | | | | | | | | | | | |
436 ValLeuGluLysGlnAlaIleGlnGlnMetValargGluTrp..... 449
  :::| | | | | | | | | | | | | | | |
1026 CCTGAGAGGAGGCTGGATGCTTCGGGGGATATATAACATTGG 1071
  :::| | | | | | | | | | | | | | | |
450 .....MetLeuAlaValSerSerAspTrp 457
  :::| | | | | | | | | | | | | | | |
seq_name: sp_fungi:Q9C105
seq_documentation_block:
ID Q9C105 PRELIMINARY; PRT; 1236 AA.
AC Q9C105;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE GLUCOAMYLASE I (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR
DE STARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAI, CONTAINS
DE CHITINASE FAMILY SIGNATURE.
GN SPAPBIE7.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL590605; CAC36921.1;
DR InterPro; IPR001917; AminoTransf_2.
DR Pfam; PF001192; chitinase_2.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
SQ SEQUENCE 1236 AA; 123387 MW; 5A2D33A30B87CDD8 CRC64;

alignment_scores:
  Quality: 127.50      Length: 359
  Ratio: 0.664         Gaps: 16
  Percent Similarity: 53.482      Percent Identity: 27.019

alignment_block:
US-09-886-400-3/rev x Q9C105 ..
Align seg 1/1 to: Q9C105 from: 1 to: 1236

1088 TCCCATTTTACCCCTC.....CAATGTTATATATCGCCCGAAGGC 1045
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
492 SerProGlnSerThrLeuSerThrSerGluValValSerGluValSe 508
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1044 ATCCAGCCCTCTCTCAGGAGGGGCTCCATCCC.....CTGCAATGCC 1001
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
508 rSerThrLeuLeuSerGlySerSerAlaIleProSerThrSerSert 525
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1000 TGTTCTCGGTAAAGGGGAGTTCGCCCTCATATTTAGGACAGACATA 951
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 hrProSerSerIleIleSerProMet.....ThrSerVal 538
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
950 TTA.....AGTCTTCGTTCCCTTCGTC.....TCTCT 922
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539 LeuSerSerSerSerIleProThrSerSerSerSerAspPheSerSe 555
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
921 CCATATCCCTCAAGCTTTATCTGTGTGGTCCCACTCGAAGTCGTAAGA 872
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
555 rSerIleThrThrIleSerSerGlyIleSer.....S 567
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
871 GTCCTCTCCACTGTCTCAGCTCTGAGGAGGACACAGTTCGAGTTG 822
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 erSerIleProSerThrPheSerSerVal.....SerSerIleLeu 580
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 AGTCGTCTATACCTCTATATATATCCCTCAACACTATCTGTAGCCTGC 772
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 SerSerSerThrSerProSerProSerSerThrSerLeuSerIleSerSe 597
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
771 AATGTCCTATAGCCATGAACCTATATATCGGTGCGGTATAGAGAATGT 722
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 rSerThrSerSerThrPheSerSerAlaSerThrSerSerProSerSerI 614
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 TGCCTCTGTCTCTATCCAGCTCGCCACTTCTTAGGATTATAGAGGA 672
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 leSerSerSerIleSerSerSerSerThrIleLeu..... 625
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
671 AGCTTCCTGATGCGAGCATTAACAGCGGTGTCAAGCCACACCAA..... 627
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
626 SerSerProThrProSerThrSerSerSerLeuMetIleSerSerSerIl 642
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
626 ....ACGGTACGGCTTCGATGCTTTGACTGCCTTTACGTTACCTTAC 581
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
642 eIleSerGlySerSerSerIleLeuSerSerSerIleSer...ThrIleP 658
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 CTTCAAAAGACGCTTATCGCTTCTTCTAGCTCCCTGAGACCAAGGAGA 531
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 rIleSerSerSerSerThrSerSerSerSerValIleProSerSer 674
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 TAGCTGATGATACCTAAAGCGCTTTTCCCTTTGGGCTTTTATAAGGTGG 481
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 .....

```

```

480 ATAGACGGCTTTAATGGCTTTATCGCGAGTGTAGATGAGCTGAGAAA 431
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
675 ....SerThrLeuValSerSerSerSerSerLeuValSerSerSerP 690
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 GCATCCCTCCCGCTCGCGAATAGATACATACCGTGTGCTTCAGT 381
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
690 roValAlaSerSerSerSerProIle.....ProSerSerSerSer 704
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
380 ATGGCAGGATTATCGGTTCATAGCGAGCTCTGGCACCCAGATCCCTT 331
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 LeuValSerThrTySerAlaSerLeuSerAsnIleThrHisSerSerLe 721
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
330 TGGAGAAAGCTCGAAGAGCTCT.....TCCTTAACCT 299
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 uSerLeuThrAlaMetSerSerSerSerAlaIleProThrSerValAsn 738
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
298 CCTATCTCTGAACCTGTGCTCT.....ACTCTGTAAGCGGAGG 255
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
738 erSerThrLeuIleThrAlaSerSerSerAsnThrLeuLeuSerSerIle 754
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 AGGGGAGTATTGCGTGTAGCTGCTCGATTATCTCTATCAGGTC 205
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
755 ThrSerSerSerAlaIleValSerSerThrThrValSerAsnIleSerSe 771
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 ACTCGGATGCCCCCTTTAACAGAGTCTATATATCCTTCGGGAGAACT 155
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
771 r...AsnLeuProSerAlaThrAlaSerSerGlnSerGlnLeuThrAsn 787
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 TTAAGCTATAGCCGTTATGTTGAGCCCAAGAAATTCTTCT...TTA 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 erSerThrLeuIleThrSerLeuTyLeuSerSerSerSerSerArgThr 803
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 ATCAGTCTCTCGATGACTGGGATGAT 81
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 IleSerThrSerSerThrAsnGluTy 812
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: sp_archaeap:Q9V294
seq_documentation_block:
ID Q9V294 PRELIMINARY; PRT; 1362 AA.
AC Q9V294;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE AMVLOPULULANASE.
GN APU OR PAB0122.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49104.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03065; Glyco_hydro_57; 1.
KW Complete proteome.
SQ SEQUENCE 1362 AA; 154578 MW; 277AFAB4E14860D1 CRC64;

alignment_scores:
  Quality: 124.00      Length: 381
  Ratio: 0.756         Gaps: 16
  Percent Similarity: 43.045      Percent Identity: 21.522

alignment_block:
US-09-886-400-3 x Q9V294 ..
Align seg 1/1 to: Q9V294 from: 1 to: 1362

```

```

94 ATCGAGACACTGATTAAAGAGAAATTCCTTTTGGCTCAACATAACGGG 143
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 ValGluThrValLeuLysHisGlnMet...TrpLeuLeuAsnHisThr... 250
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
144 CTATACCTTAAAGTTCTCCGGAAGGATATATAGACCTCGTTAAAGGG 193
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
251 .....PheGluHisGluLysIleAsnLeuLeuLeuGly. 262
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
194 GCATCGCGAGTGCCTCATAGACATATCGGAACGAGCTACACGACGCA 243
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
263 .....AsnGlyAsnValGluValThrValValProTyrThrHisPro 276
244 ATACTCCCTCTCTCCG.....CTTAGCAGAGTAGAGCAACA 281
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
277 IleGlyProIleLeuAsnAspPheGlyTyrTrpTyrGluAspPheAsp 293
282 AGTTCAGAGAGTAGGAGCTTAAGGAAGAGCTCTTC.....GAGC 322
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
293 nValLysLysAlaAsnGluLeuTyrLysGluTyrLeuGlyAlaGlyLysV 310
323 TTCTCCAAAGGATTCGTGCTGCCAGAGCTCGCCTATGACCCGATAATC 372
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310 alThrProLysGlyGlyTyrAlaAlaGluSerAlaLeuAsnAspLysThr 326
373 CCTCCACTAGAGGACACGCTTATGACTATCTATCGCCGACGGG... 420
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
327 LeuGluLeuLeuAlaAsnGlyTyrLysIleValMetThrAspGlnLe 343
421 .....GAGCGCATGCTTTCTCAGCTCATCTCACTCGCGGATAAAGC 463
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
343 uValLeuGluLysLeuGlyValProLysThrIleGluSerTyrTyrLysP 360
464 CA..... 465
360 rotTrpValAlaGlnPheGlyAspLysLysIleTyrLeuPheProArgAsn 376
465 ..... 466
377 HisAspLeuSerAspArgValGlyPheArgTyrAlaGlyMetAsnGlnTyr 393
466 .....ATTAACCGCTCTATCCACACCTTATAAGGCCCAAGGGAAA 508
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
393 rAspAlaValLysAsnPheValGluGluLeuLeuLysIleGlnLysGlnA 410
509 AGCGCTTTAGTACATCAGCTATCTCTCTGCTCTCAGGGAGCTTAGGAAG 558
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
410 snTyrAspGlySerLeuValTyrValIleThrLeuAspGlyGluAsnPro 426
559 GCGATAAGCTCGTTTGAAGGTAAAGTAAACGCTAAAGSCAGTC...AA 605
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
427 TrpGluHisTyrProPheAspGlyLysLeuPheLeuGluLeuTyrTr 443
606 AGACATCGAACCCCTACCCGTTTGGGTGGCGCTGAACACGGCTGTA.... 651
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
443 gGlnLeuGluLeuGlnLysLysGlyLeuIleArgThrValThrProS 460
652 .....ATGCTCGGC.....ATCGGAAGGCTTCCTCTTATG 681
460 erGluTyrIleGluMetPheGlyAspLysAlaAsnLysLeuThrProLys 476
682 AATCCTAAGAAGTGGGAGCTGTGATAGAGCAAGGACAACATCTTCT 731
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
477 MetMetLysArgLeuAspPheThrThrGluAspAsnValAsnAlaLeu 493
732 ATAGGCAACCATATAGAGTTTCATTGGCTATAGGACATTGCGAGGTACA 781
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
493 uLysAlaLysThrLeu.....GlyGluL 501
782 GAATGAGTGTGAGGATATATAGAGTTATAGACGAGCTCAACTCGGAA 831
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 eutyAspMetValClyValThrGlu..... 509

```

seq\_name: sp\_archaeap:Q9HL91

seq\_documentation\_block:

ID	Q9HL91	PRELIMINARY;	PRT;	357 AA.
AC	Q9HL91;			
DT	01-WAR-2001 (TREMELrel. 16, Created)			
DT	01-WAR-2001 (TREMELrel. 16, Last sequence update)			
DT	01-WAR-2001 (TREMELrel. 16, Last annotation update)			
DE	ALPHA-AMYLASE RELATED PROTEIN.			
GN	TA0339.			
OS	Thermoplasma acidophilum.			
OC	Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;			
OC	Thermoplasma			
OX	KCBI_TaxID=2303;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DSM 1728;			
RX	MEDLINE=20479972; PubMed=11029001;			
RA	Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,			
RA	Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;			
RT	"The genome sequence of the thermoacidophilic scavenger Thermoplasma			
RT	acidophilum."			
RL	Nature 407:508-513(2000).			
DR	EMBL; AL445064; CAC11483.1; ..			
KW	Complete proteome.			
SQ	SEQUENCE 357 AA; 41845 MW; B8C1A97F4EB237F4 CRC64;			

alignment\_scores:

Quality	122.50	Length:	307
Ratio:	0.869	Gaps:	8
Percent Similarity:	45.928	Percent Identity:	16.612

alignment\_block:

US-09-886-400-3 x Q9HL91

Align seg 1/1 to: Q9HL91 from: 1 to: 357

64 AAGGTCATAGAGAAGGATACATCCAGTCATCGAGACACTGATTAAAGA 113

20 ArgIleAlaAlaAsnAsnTyrIleProAlaThrGlnSerLeuMetAsp 36

114 AGAATATCCITTTGGGCTCAACATAACGGGCTATACCTTA..... 153

36 rGlyIleArgSerSerPheSerLeuSerGlyThrPheMetGluGlnAla 53

154 ..AAGTTCCTCCGAAGGATATTATAGACCTCGTTAAAGGGGCGCATCGG 201

53 erLysTyrCysProLys...ValIleAspValIleAspTyrValArg 68

202 AGTACCTGATAGAGATATCGGAACGAGCTACACGACCACTACTCC 251

69 SerGlyGlnCysGluLeuLeuSerGluThrTyrHisSerLeuAla 85

```

252 CTTCTCCCGCTTACGAGAGTAGAAGCACCAAGTTCAGAGAGATAGGGAAG 301
   ::::::::::: |||||::: :::::::::::
85 rIleTrpAsnAspGluGluPheValArgGlnValArgMetGlnGluSerA 102
   ::::::::::: |||||::: :::::::::::
302 TTAAGGAAGAGCTTCGAGCTTCTCCAAAGGGATTCTGGCTGCCAGAG 351
   ::::::::::: |||||::: |||||::: |||||:::
102 latLeuLysThrPheAsnTyrGluProValSerPheArgAsnThrGlu 118
   ::::::::::: |||||::: |||||::: |||||:::
352 CTGCGCTATGACCGGATATCTCCCTGCCATCTAGGAAGGACACGGTTATGA 401
   |||||::: |||||::: |||||::: |||||:::
119 LeuIleTyrAsnAspHisIleAlaGluValAlaLysGlyMetGlyPheAr 135
   ::::::::::: |||||::: |||||::: |||||:::
402 GATCATCTTCGCGAGGGGAGCGATGCTTTTCACGCTCATCTCAACT 451
   ::::::::::: |||||::: |||||::: |||||:::
135 gAsnIleLeuAlaGluGlyThrAspAspIleAlaSerArgTyr..... 149
   ::::::::::: |||||::: |||||::: |||||:::
452 CGCGCGATAAGCAATTAACCGCTCTATCCACACCTTATAAGGCCCAA 501
   ::::::::::: |||||::: |||||::: |||||:::
149 ..... 149
502 AGGGAAGCGGTTAGGTAC.....ATCAGCTATCTCT 536
   ::::::::::: |||||::: |||||::: |||||:::
150 ...AspValAsnTyrArgTyrAlaAlaProSerGlyIleAsnLeuTyrLe 165
   ::::::::::: |||||::: |||||::: |||||:::
537 TGGTCTCAGGAGCTTAGGAAGCGGATAAGCTCGTTTGAAGGTAAGG 586
   |||||::: |||||::: |||||::: |||||:::
165 uArgAsnTyrProLeuSerAspAspIleSerPheArgPheSerAsnArgA 182
   ::::::::::: |||||::: |||||::: |||||:::
587 TAACGCTAAGGAGCTCAAGACATCGAAGCGGTACCCGTTTGGGTGCC 636
   ::::::::::: |||||::: |||||::: |||||:::
182 latTrpAlaAspTyrProLeuThrAlaAspLysPheAlaArgTrpIleSer 198
   ::::::::::: |||||::: |||||::: |||||:::
637 GTGACGAGCGGTATGCTCGGCATCGGAGGCTTCCTCTATGAAATCC 686
   ::::::::::: |||||::: |||||::: |||||:::
199 AlaSerSerGlyIleValAsnLeu..... 207
   ::::::::::: |||||::: |||||::: |||||:::
687 TAAGAAAGTGGGAGCTGGATAGAGACAAAGCAACATTTCTTATACG 736
   ::::::::::: |||||::: |||||::: |||||:::
207 ..... 207
737 GCACCGATATAGATTTCATGGCTATAGGACATTCGAGGCTACAGATG 786
   ::::::::::: |||||::: |||||::: |||||:::
208 .....PheMetAspTyrGluThrPheGlyGluHisGlnArg 219
   ::::::::::: |||||::: |||||::: |||||:::
787 AGTGTGAGGATTATAGAGTTATA..... 813
   |||||::: |||||::: |||||::: |||||:::
220 ProGluThrGlyIlePheGluPheLeuArgTyrLeuProMetTyrPheG1 236
   ::::::::::: |||||::: |||||::: |||||:::
814 ....GACGAGCTCACTCGAAGCTGTGCTTCCTCCCTCAGAGCTGAAGCACA 859
   ::::::::::: |||||::: |||||::: |||||:::
236 uGluAsnAspValHisThrIleLeuValArgGluAlaGluAlaArgHisA 253
   ::::::::::: |||||::: |||||::: |||||:::
860 GTGGAAGGGAGCTC.....TACTTACGAGCTTCGAGTTGGSCA...CCA 900
   |||||::: |||||::: |||||::: |||||:::
253 rgThrArgAspPheIleSerValSerLysThrThrSerTrpAlaAspLys 269
   ::::::::::: |||||::: |||||::: |||||:::
901 GATAAGAGCTTGAGGTATCG 921
   ::::::::::: |||||::: |||||::: |||||:::
270 AsnArgAspLeuSerAlaTrp 276
   ::::::::::: |||||::: |||||::: |||||:::
seq_name: sp_archaeap:030246
seq_documentation_block:
ID O30246 PRELIMINARY; PRT; 324 AA.
AC O30246;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE METHANOL DEHYDROGENASE REGULATORY PROTEIN (MOXR).
GN AF2425.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

```

```

OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Raine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:384-370(1997).
DR EMBL; AE001108; AAB91247.1; -.
DR TIGR; AF2425; -.
DR InterPro; IPR000523; Mg_chelatase_chiI.
DR Pfam; PF01078; Mg_chelatase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 324 AA; 36609 MW; B0F64965EFC82F1F CRC64;

```

## alignment\_scores:

Quality: 120.50 Length: 327  
Ratio: 0.748 Gaps: 14  
Percent Similarity: 49.235 Percent Identity: 21.713

## alignment\_block:

US-09-886-400-3 x O30246

Align seg 1/1 to: O30246 from: 1 to: 324

```

1 TTGAGAGCGCTGCTCTTTTCACGGCAACCTCCAGTATGCCGAATCCCA... 48
   ||| ||| :::: |||||::: :::: |||
35 LeuAlaAlaLeuThrAsnGlyAsnIleLeuPheGluAspTyrProG1 51
49 .....AGAGCGAAATCCCAAGGTCATAGAGAGGCATACATCCAG 91
   |||||::: :::: |||||::: :::: |||
51 yLeuGlyLysThrLeuLeuAlaLysValPheAlaArgVal..... 64
92 TCATCAGACACTGATTAAGAAAGAAATTCCTTTTGGCTCAACATAACG 141
   |||||::: |||||::: |||||::: |||
65 .....IleGlyAlaAspTyrArgVal 72
142 GGTATACCTTAAGTCTCTCCGGAAGATATTATAGACCTCGTTAAGG 191
   ::|||::: |||||::: |||||::: |||||
73 GlnPheThrProAspLeuLeuProSerAspIleIle..... 84
192 GGCATC.....GCGAGTCACTGATAGATAATC...GGAAGA 229
   |||||::: :::: |||||::: |||||::: |||
85 .GlyValLysIleTyrArgGlyAspArgPheGluPheValLysGlyProI 101
230 GTTACACGCACTACTC.....CCCTC 255
   ::|||::: |||||::: |||||::: |||
101 lePheThrAsnValLeuLeuAlaAspGluIleAsnArgSerProProlys 117
256 CTCCTGCTTAGCAGAGTAGAAGCACAAAGTTCAGAGATAGGAAGTTAA 305
   :::: |||||::: |||||::: |||||::: |||
118 ThrGlnAlaLeuLeuGluAlaMetGluGluLysGlnIleThrValG1 134
306 GGAAGAGCTCTTCGAGCTTCTCCAAAGGATTCGTGCTGCCAGAGTCG 355
   ::|||::: |||||::: |||||::: |||
134 uGlyGluThrPheSerLeuSerMetProPhePheValLeu.....Alat 149
356 CCTATGACCGGATATCCCTGCCATCTCAAGACACGGTTATGATAT 405
   :::: |||||::: |||||::: |||||::: |||
149 hrGlnAsnProfile..... 153

```

```

406 CTATTCGCCGACGGGAGGCGATGCTTTCTCAGCTCATCTCAACTCGC 455
154 ...GluGlnGluGlyThrProLeuProGluAlaGlnMetAspArgPh 169
456 GATAAGCCCAATTAAGCGCTCTATCACACCTTATAAAGGCCCAAGG 505
159 eMetLeuArgMetArgProGlyThrProGluSerIleGluGluMetG 186
506 AAAAGCGCTTTAGTACATCAGCTAT ..... 531
186 LuIleLeuArgArgArgIleSerTirpArgLysAspAspProThrGluAsp 202
532 .....CTCCTTGCTCTCAGGAGCTTAGGAGCGAT 563
203 ValGluProValValSerLeuGluThrPheArgArgIleGlnAspAla 219
564 AAAGCTGCTTTTGAAGTAGTAAAGCTTAAGCGGAGTCAAGACATCG 613
219 lGluAlaValTyValAspLysSerIleLeuLysTyIleSerGluLeuV 236
614 AAGCGGTACCGCTTGGTGGCGCTGAAACAGCGCTGATATGCTCGGCATC 663
236 alArgAla .....ThrArgGluHisGluLeuValGluLeuGlySer 249
664 .....GGAAGGCTCCTCTTATGAATCCTAAGAAAGTGGCGAGCTG 704
250 SerProArgGlyGlyLeuAlaLeuLysLeuAlaArgAlaLeuAlaV 266
705 GATAGAGACAAAGCAACATCTT .....CTATCGGACCCGATA 745
266 lMetAspGlyArgAspPheValIleProAspAspValLysArgValAlaV 283
746 TAGACTTCATTCGCTATAGGACATTCAGCGCTACAGATGAGTGTTGAG 795
283 alGluAlaLeuAlaHisArgValIleLeuLysPheGluTyAlaValGlu 299
796 GGATATTA .....GAGGTATAGACAGCTCAACTCGGAAGTGCCT 839
300 GlyLeuArgAlaGluGluValGluAlaGluGluLeuLeuAsnSerValArgVa 316
840 TCCCTCAGAGCTGAGCACACGGAAGGGAG 870
316 lPro .....LysTyArgAlaGlnGlu 323

seq_name: sp_archaeap:Q9V0M7

seq_documentation_block:
ID Q9V0M7 PRELIMINARY; PRT; 602 AA.
AC Q9V0M7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 70.2 KDA PROTEIN.
GN PAB1857.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=23292;
FN [1]
RC STRAIN=ORSAY;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; CAB49676.1; -
DR InterPro; IPR001395; Aldo_ket_red.
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
DR SMART; SM00278; HH1; 1.
DR PROSITE; PS00062; ALDOKETC_REDUCTASE_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 602 AA: 70219 MW; 326B1198D54DD43E CRC64;

alignment_scores:
Quality: 119.00 Length: 157
Ratio: 1.545 Gaps: 4
Percent Similarity: 49.045 Percent Identity: 26.752

alignment_block:
US-09-886-400-3 x Q9V0M7
Align seg 1/1 to: Q9V0M7 from: 1 to: 602

40 GAATATCCCAAGACGCGAAATCCCAAGAGCTCATAGAGAAGCATACATCCC 89
71 GluTyrlleysAlaGluPheAsnArgTyrlMetGluArgLysle..... 85
90 AGTCATCGACGACACTGATTAAAGAAGAAATTCCTTTGGCTCAACATAA 139
86 .....LysLeuMetLysGluAspLeuLysLysAlaAspGlyLysL 99
140 CGGCTATACCTTAAAGTTCCTCCCGAAG..... 168
99 euArgAsnAlaIleGluPheMetIleLysTyrlPheGluAspValTyrlGlu 115
169 .....GATATTATAGACTCGTTAAAGGGGGCAT 197
116 TyrlPsrLysIleAsnGlyAspIleIleGlyArgPheLysGlnLeuGl 132
198 CGCAGAGTACCTGATAGATATATCGGAAGAGCTACACGCGCAATAC 247
132 nAspGluGlyPheValGluIleIleThrSerAlaAlaThrHisGlyTyrl 149
248 TCCCGCTCCCTCCCGCTTAGCAGA...GTAGAAGCACAAAGTTCAGAGAT 294
149 euProLeuLeuGlyArgAspGluAlaIleAspAlaGlnIleLeuThrGly 165
295 AGGAAGTAAAGGAAGAGCTTCGAGCTTCTCCAAAGGATTCGTGGCT 344
166 lIleArgValTyrlGluLysTyrlPheGlyLysLysProArgLyleIleTrpLe 182
345 GCCAGAGCTCGCTATGACCGATA..... 369
182 uProGluCysAlaTyrlArgProAspGlyLeuTyrlLysSerProSerThrG 199
370 .....ATCGCTGCCCATCTAGGAGACACGCT 396
199 LyGluIleLysTyrlArgLysGlyIleGluHisPheLeuLysLysTyrlGly 215
397 TATGAGTATCTATTCGCGGAC 417
216 LeuGluPhePheValGlu 222

seq_name: sp_bacteriap:083377

seq_documentation_block:
ID 083377 PRELIMINARY; PRT; 526 AA.
AC 083377;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0358.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
FN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=NICHOLS;
RC MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

```







```
|||||
134 luTyLeuPheAlaAspGlyGluAlaMetLeuPheSerAsnHisLeuAsn 150
451 TCGCGATAAGCAATTAACACGCTATCCACACCTTATTAAGGCCCA 500
|||||
151 SerAlaIleLysSerIleLysProLeuTyProTyLeuIleLysAlaG 167
501 AAGGAAAGCGCTTTAGGTACATACAGTATCTCTTGGTCTCAGGGAGC 550
|||||
167 nArgGlyGluGlyPheValTyLeuAsnTyLeuLeuGlyLeuArgGlu 184
551 TTAGGAAGCGGATAAAGCTGCTTTTGAAGCTAAGCTAAGCTAAGGCA 600
|||||
184 euLysLysAlaIleAsnLeuThPheGlyGlyLysValThrLeuGluAla 200
601 GTCAAGACATCGAAGCGCTACCGCTTGGTGGCGGTGAACAGCGCTGT 650
|||||
201 ValLysAspIleGluAlaIleProValTrpValSerIleAsnIleAla 217
651 AATGCTGGCATCGAAGGCTTCTCTATGAATCCTTAAGAAAGTGGCGA 700
|||||
217 eMetLeuGlyAlaGlyArgPheProLeuMetSerProLysArgValAla 234
701 CTGATAGAGGACAAGACACATCTTCTATACGCCACCGATATAGAG 750
|||||
234 snTrpIleLysGlyLysAspGluIleLeuLeuTyrglyThrAspIleGlu 250
751 TTCATTGGCTATAGGACATTCAGGCTACAGATGAGTGTGGAGGATT 800
|||||
251 PheLeuGlyTyArgSerIleAlaGlyHisLysIleAlaIleSerGlyLe 267
801 ATTAGAGTTATAGACAGCTCAACTCGGAAGCTGCTTCCCTCCCTCAG 850
|||||
267 uMetGluValLeuAsnGluLeuGlyGlyLeuLeuCysLeuProArgAsp 284
851 TGAAGCAGTGGGAGGAGCTCTACTACGAGCTTCGAGTGGGACCA 900
|||||
294 leArgHisAsnGlyArgLeuTyLeuArgThrSerTrpAlaPro 300
901 GATAAGCTTTGAGGATATGAGAGAGGAGGAGGAGGAGGAGGAGGAG 950
|||||
301 AspLysSerLeuArgIleTrpLysGluAspGluGlyAsnAlaArgLeu 317
951 TATGCTGCTACAAATATAGGCGGAGCTCGCCCTTTAGCGGAGAAC 1000
|||||
317 nMetLeuThTyrcysMetAspGlyGluPheAlaPheLeuAlaGluAs 334
1001 CGGATGCAAGGGGATGGAGCCCTCCCTCAGAGAGGCTGGATGCTTC 1050
|||||
334 erAspAlaArgGlyTrpGluProLeuProGluArgArgLeuAspAlaPhe 350
1051 CGGCGATATATACGATTGGAGGGTGAATGGGAA 1089
|||||
351 LysAlaIleTyLysTyTrpArgAsnGluAsnGlyLys 363

seq_name: pir2:AC2112
seq_documentation_block:
hypothetical protein alr2450 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #tax_change 11-Jan-2002
C:Accession: AC2112
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2112
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB74149.1; PID:gl7131542; GSPDB:GN00179
```

```
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2450
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

alignment_scores:
Quality: 169.00 Length: 495
Ratio: 0.820 Gaps: 21
Percent Similarity: 41.616 Percent Identity: 21.010

alignment_block:
US-09-886-400-3 x AC2112
Align seg 1/1 to: AC2112 from: 1 to: 529

7 CGCTCTCTTTTACGGCACCTCCAGTATGCGGAAATCCCAAGAGCGA 56
|||||
7 AlaleuValleuHisAlaHisLeuProPheValArgHisProGluSerAs 23
57 A.....ATCCAAAGTCATAGAGAAGGCATACA 85
:|||||
23 pTyrvAlleuGluGluLutrpLeuTyrgluAlaIleThrGluThrTyrl 40
86 TCCCA.....GTCATCGAGACACTGATTAAAGAGAATTCCTTTT 126
|||||
40 leProLeuLeuLysValPheAspGlyLeuLysArgAspGlyIleAspPhe 56
127 GGGCTC..... 132
:
57 LysIleThrMetSerMetThrProProLeuValSerMetLeuArgAspPr 73
132 ..... 132
73 oLeuLeuGlnGluArgTyrgluAlaHisLeuAlaGlnLeuGluGluLeu 90
133 .....AACATACGGGCTATACCTTAAAGTTCCCTC 162
|||||
90 leGluLeuGluGlyGluArgAsnAlaGlnAsnGlyHisLeuArgTyrglu 106
163 CCGAAG..... 168
:
107 AlaGluHisTyrgluAlaThrGluPheAsnGluAlaArgGlnMetTrpGlu 123
169 .....GATATTATAGACCTCTTAAAGGGGCATCGCGAGTGACC 208
|||||
123 gTyrgAsnGlyAspLeuValThrAlaPheLysGlnPheGlnAspSerAsn 140
209 TGATAGAGATAATCGGAACGAGCTACACGACCAATACCTCCCTCCCTC 258
|||||
140 snLeuGluIleIleThrCysGlyAlaThrHisGlyTyrgluProLeuMet 156
259 CCGCTT.....AGCAGATAGACACAGTTCAGAGATAGGAAGT 302
:
157 LysMetTyrgProGlnAlaValTrpAlaGlnIleGlnValAlaCysGluH 173
303 TAAGGAAGAGCTCTTCGAGCTTTCTCCAAAGGATTCCTGCTGCCAGAGC 352
|||||
173 sTyrgGluLutrpPheGlyArgProProLysGlyIleTrpLeuProGluC 190
353 TCGCTATGACCGGATATCCCTCCCATACTGAGGACACAGGTTATGAG 402
|||||
190 ysAlaTyrgTyrgluGlyLeuGluArgMetLeuAlaAspAlaGlyLeuArg 206
403 TATCTATTCCGCGAGGGGAGGATGCTTTTCTCAGCTCATCTCAACTC 452
|||||
207 TyrgPheLeuThrAspGlyHisGlyIleLeuTyrg.....AlaArgProArgPr 222
453 GCGGATAAAGCCAAATTAAACCGCTC..... 477
:
222 oArgPheGlyThrTyrgAlaProIlePheThrGluThrglyValAlaAlaP 239
477 ..... 477
```

239 heGlyArgAspHisGluSerSerGlnGlnValTrpSerSerGluValGly 255  
 478 TATCCACACCTTATAAAGGCCCAAGG.....GA 506  
 ||||| : : : : : ||  
 256 TyrProGlyAlaAlaGluTyrArgGluPheTyrLysAspLeuGlyTrpG1 272  
 ||||| : : : : : ||  
 507 AAAGCGCTTAGGTACATCAGCTATCTCTTGGTCTCAGGGAGCTTAGGA 556  
 || : : : : : ||||| : : : : : |||||  
 272 ualaGluTyrGluTyrLysProTyrIleMetProAsnGlyGlnArgL 289  
 || : : : : : ||||| : : : : : |||||  
 557 AGCGGATAAAGCTCGTTTGAAGTAAGTAACG..... 591  
 || : : : : : ||||| : : : : : |||||  
 289 ysAsnThrGlyIleLysTyrHis...LysIleThrGlyArgGlyLeuGly 304  
 592 .....CTAAGGCGAGTCAAGACATCGAAGCCGTACCCGTTTGGGTGGC 635  
 ||||| : : : : : ||||| : : : : : |||||  
 305 LeuSerAspLysAlaLeuTyrAsp.....ProTyrTrpAlaLys 317  
 636 CGTGAACACGCTGTAATGCTCGGCATCGGAAGCTTCCTCTTATGAATC 685  
 : : : : : ||||| : : : : : |||||  
 317 sGluLysAlaAla..... 321  
 686 CTAAGAAATGCGAGTGGATAGAGACAGGACAAACATCTT..... 729  
 : : : : : ||||| : : : : : |||||  
 322 ..GluHisAlaAlaAsnPheMetTyrAsnArgGluArgGlnAlaGluHis 337  
 ||||| : : : : : ||||| : : : : : |||||  
 730 CTATACGCGCAC.....GATAT 746  
 ||||| : : : : : ||||| : : : : : |||||  
 338 LeuTyrGlyIleMetGlnArgProProIleValSerProTyrAspAl 354  
 ||||| : : : : : ||||| : : : : : |||||  
 747 AGAGTTCATTGGCTATAGGACATTGCAGG..... 777  
 ||||| : : : : : ||||| : : : : : |||||  
 354 aGluLeuPheGlyHisTrpTrpTyrGluGlyProTyrPheIleAspTyrL 371  
 ||||| : : : : : ||||| : : : : : |||||  
 778 .....TACAGATAGTGTGGAGATTATTAGAGTT 810  
 ||||| : : : : : ||||| : : : : : |||||  
 371 euPheArgLysSerTyrTrpAspGlnGlyThrTyrAlaMetThrHisLeu 387  
 ||||| : : : : : ||||| : : : : : |||||  
 811 ATAGACGAGCTCAACTCGAA.....CTGTGCTTCCTCCACAGA 848  
 ||||| : : : : : ||||| : : : : : |||||  
 388 AlaAspTyrLeuArgAsnGluProThrGlnValCysArgProSerGI 404  
 ||||| : : : : : ||||| : : : : : |||||  
 849 GCTGAAGCACAGTGGAGGGAGCTTACTTACGAGCTTCGAGTTGGGCAC 898  
 : : : : : ||||| : : : : : |||||  
 404 n.....:SerSerTyrpLyt 409  
 899 CAGATAAGAGCTTGAGGATATGAGAGAGGAGGAGGAAAGCA..... 942  
 : : : : : ||||| : : : : : |||||  
 409 yLysGlyPheHisGluTyrTrpLeu...AsnGluThrAsnAlaTrpIle 424  
 943 .....AGACTTAATATCTCTCTACAAATATAGGGGCGAATCGCCCT 986  
 : : : : : ||||| : : : : : |||||  
 425 TyrProHisLeuHisLysAlaAlaGluArgMetile...GluIleSerTh 440  
 ||||| : : : : : ||||| : : : : : |||||  
 987 TTTACCCGAGAACACCGATGCAAGGGATGGAGCCCTCCCTCAGAGGA 1036  
 ||||| : : : : : ||||| : : : : : |||||  
 440 rLeu...GluProGluAspGluLeuGlyTrpArgAlaLeuAsnGlnAla 456  
 1037 GGCTGGATGCGCTTCGGCGCATATATACGATTGG 1071  
 : : : : : ||||| : : : : : |||||  
 456 laArgGluLeuLeuAlaGlnSerSerAspTrp 467

seq\_name: pir2:S76831

seq\_documentation\_block:

hypothetical protein - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S76831

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis.

A:Reference number: S74322; MUID:97061201

A:Accession: S76831

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 &lt;RAN&gt;

A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAAL8743.1; PID:g16  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

alignment\_scores:

Quality: 162.50 Length: 471

Ratio: 0.860 Gaps: 19

Percent Similarity: 40.127 Percent Identity: 21.444

alignment\_block:

US-09-886-400-3 x S76831 ..

Align seg 1/1 to: S76831 from: 1 to: 529

7 GCGCTGCTCTTCACGGCACACTCCAGTATCCGAAATCCCAAGACGGA 56

||||| : : : : : ||||| : : : : : |||||

7 AlaLeuValLeuHisAlaHisLeuProPheValArgHisProGluSerAs 23

57 A.....ATCCCAAGGTCATAGAGAAGGCATACA 85

: : : : : ||||| : : : : : |||||

23 pTyrValLeuGluGluTrpLeuTyrGluAlaIleThrGluThrTyrI 40

86 TCCAGTCATC.....GAGACACTGATTAAAGAAAGAAATTCCTTTT 126

||||| : : : : : ||||| : : : : : |||||

40 leProLeuIleHisValPheGluGlyLeuLysArgAspGlyValAspPhe 56

127 GGGCTCAACATAACG..... 141

: : : : : ||||| : : : : : |||||

57 LysIleThrMetSerMetThrProProLeuValSerMetLeuArgAspPr 73

141 ..... 141

73 cLeuLeuGlnGlnArgTyrGluAlaHisLeuSerLeuLeuGlnGluLeu 90

142 .....GGCTATACC 150

90 euAlaLysGluIleValArgAsnGluHisAsnGlyHisLeuGlnTyrLeu 106

151 TTAAGTTCTCCCGAAG..... 168

||||| : : : : : ||||| : : : : : |||||

107 AlaAspPheTyrAlaLysGluPheAlaIleArgGluThrTrpGluAr 123

169 .....GATATTATAGACCTCGTTAAAGGGGCATCGGAGTGACC 208

||||| : : : : : ||||| : : : : : |||||

123 gTyrAspGlyAspLeuValThrAlaPheLysGlnPheGlnAspSerAsn 140

209 TGATAGACATAATCGGACGAGCTACAGCAGCAATACTCCGCCCTCCTC 258

||||| : : : : : ||||| : : : : : |||||

140 snLeuGluIleIleThrCysGlyAlaThrHisGlyTyrPheProLeuMet 156

259 CGCTT.....AGCAGATAGAACACAAAGTTTCAGAGAGATAGGGAAGT 302

: : : : : ||||| : : : : : |||||

157 LysMetTyrProGlnAlaValTrpAlaGlnIleLysValAlaCysGluH 173

303 TAAGGAAGAGCTCTTCGAGCTTCTCCAAAGGATTCTGGCTGCCAGAGC 352

||||| : : : : : ||||| : : : : : |||||

173 sTyrGluGluAsnPheGlyArgSerProLysGlyIleIleProGluC 190

353 TCGCCTATGACCCGATAATCCCTGCCCATCTGAAGGACAAACGGTTATGAG 402

||||| : : : : : ||||| : : : : : |||||

190 ysAlaTyrTyrGluGlyValGluArgMetLeuAlaAspAlaGlyLeuArg 206

403 TATCTATTTCGGCAGCGGAGGCGCATGCTTTCTCAGCT..... 441

||||| : : : : : ||||| : : : : : |||||



175 heProIleAspGluLysLeuArgTyrLeuIleProPheArgProValAsn 191  
556 AAGCGATAAAGCTCGTGTTTT.....GAAGCTAAGGT 587  
192 GluThrLeuGluTyrLeuHisSerLeuAlaAspGluAspGluSerLysVa 208  
588 AACGCTAAAGGCAGCTCAAGACATCGAAGCCGTACCGTTGGTGGCCG 637  
208 lAlaVal...PheHisAspAspGlyGluLysPheGlyAlaTrpProGlyT 224  
638 TGAACACGGCTGTAATGCTCGGCATCGGAAGCTTCCTCTTATGAATCCT 687  
224 hrHisGluLeuValTyr..... 229  
688 AAGAAGTGGCGAGCTGGATA.....GA 710  
230 .....GluArgGlyTrpLeuLysGluPhePheAspArgIleSerSerAs 244  
711 GGACAAGGACAACAATTCTTATACGGCACCGATATAGAGTTTCATTGC. 759  
244 pAspLysIleAsnLeuMetLeuTyrSer.....GluTyrLeuSerL 258  
760 .....TATAGGCACATTGCAGGCTAC...AGA 783  
258 ysPheArgProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGlu 274  
784 ATGAGTGTGAGGGATTA.....TTAGAGGT 809  
275 MetSerGluTrpSerLeuProAlaArgGlnAlaLysLeuPhePheGluPh 291  
810 TATAGCAGCGCTCAACTCGGAACGTGCCCTCCCTCAGAGCTGAAGCACA 859  
291 eIleLysLysLeu...LysGluLeuAsnLeuPheGluLysTyrArg... 305  
860 GTGAAGGGAGCTCTACTTACGGACTTCGAGTTGGCCACAGATAAAGAC 909  
306 .....IlePheValArgGlyGlyIleTrp.....LysAsn 315  
910 TTGAGGATATGGACAGAGGAGGAAGAAC.....GCAACACATTAA 950  
316 Phe...LeuTyrLysTyrProGluGlyAsnTyrMetHisLysArgMetLe 331  
951 TATGCTCTCTTACAATATGAGGGCGCAACTGCCTCTTTTAGCCGAGAACA 1000  
331 uMetLeuSer.....LysLeuLeuArgAsnAsn. 340  
1001 GCGATGAAGGGGATGGAGCCCTCCCTGAGAGGAGCTGGATGCCTTC 1050  
341 .....ProThrAlaArgIlePheValLeu 348  
1051 CGGCGCATATATACGAT.....TGGAGGGGTGAAATGGG 1086  
349 ArgAlaGlnCysAsnAspAlaTyrTrpHisGlyValPheGly 362

seq\_name: pir1:ALDYAT

seq\_documentation\_block:  
 amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum  
 C:Species: Dictyoglomus thermophilum  
 C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 18-Jun-1999  
 C:Accession: S00628; A34969  
 E:Fukusumi, S.; Kamizono, A.; Horinouchi, S.; Beppu, T.  
 Eur. J. Biochem. 174, 13-21, 1988  
 A:Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic  
 A:Reference number: S00628; MUID:88225097  
 A:Accession: S00628  
 A:Molecule type: DNA  
 A:Residues: 1-686 <FUK>  
 A:Cross-references: EMBL:X07896; NID:g2688; PIDN:CAA30735.1; PID:g2689  
 A:Accession: A34969  
 A:Molecule type: protein  
 A:Residues: 2-13 <FUK2>  
 C:Genetics:

```

200 sPLysPheAlaSerGluAspLysSerLysLysLeuAlaLeuPheAsp 216
|||||
745 ATAGAGTTCATTGGC.....TATAGGCACATTGCAGGCTA 779
|||||
217 GlyLysPheGlyLeuTyrProAspThrTyrArgThrVal.....Ty 231
|||||
780 CAGAAATGAGTGTGAGGGATTATAGAGTT...ATAGACGAGCTCAACT 826
|||||
231 rGlu.....GluGlyTrpLeuGluThrPheValSerLysIleLysG 245
|||||
827 CGGAACGTGCTCCCTCA.....GAG 849
|||||
245 LuAsnPheLeuValThrProValAsnLeuTyrThrTyrMetGlnArg 261
|||||
850 CTGAAGCACAGTGAAGGGAGCTCTACTTACGAGCTTCGAGTTGG 894
|||||
262 ValLysProLysGlyArg...IleTyrLeuProThrAlaSerTyr 275
|||||
seq_name: pir2:F97197

```

```

seq_documentation_block:
uncharacterized conserved protein, related to alpha-amylase/alpha-mannosidase CAC2414 [1
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F97197
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4833-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80369.1; PID:g15025430; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

```

```

alignment_scores:
Quality: 147.00 Length: 496
Ratio: 0.766 Gaps: 20
Percent Similarity: 39.506 Percent Identity: 18.930

```

```
alignment_block:
US-09-886-400-3 x F97197 ..

```

```
Align seg 1/1 to: F97197 from: 1 to: 527

```

```

13 GTCTTTCACGGCAACCTCCAGTATGCCGAATCCCAAGAGCGAA..... 57
|||||
9 ValLeuHisSerHisMetProPheValArgHisProGluThrArgAspSe 25
58 .....ATCCCAAGCTCATAGAGGCATACATCCAG 91
|||||
25 rLeuGluGluArgTrpLeuPheGluAlaMetSerGluCysTyrIleProL 42
|||||
92 TCATCGAG.....ACACTGATTAAAGAAATTCCTTTGGG... 129
|||||
42 euIleGluValTyrAspAsnLeuLeuLysAspAsnIleLysPheArgMet 58
|||||
129 ..... 129
59 ThrMetSerIleThrProProLeuMetSerMetLeuGlnAspGluTyrLe 75
130 .....CTCAACATAACGGGCTATACCTTAAAGTTCCTCCCGA 166
|||||
75 LuAsnSerArgTyrLeuAsnTyrLeuLysLysThrIleGluLeuSerGluL 92
|||||
167 AGGTATTATA..... 177

```

```

|||||
92 ysGluIleLeuArgThrLysAsnAsnArgGluGluAsnLysValAlaLeu 108
177 ..... 177
109 PheTyrAsnLysArgAlaGluAsnThrLeuLysIleTyrGluLysTyrAs 125
178 ....GACCTCGTTAAAGGGGCATCCGAGTGAC.....CTGATAG 214
|||||
125 PaAsnAsnLeuIleAsnAlaPheArgLysTyrAspArgLeuGlyCysValG 142
215 AGATAATCGGAACGAGTACACGACGCAATACTCCCTCCCTCCCGCTT 264
|||||
142 luIleIleThrCysAlaIaThrHisAlaLeuLeuProLeuIleLeuIle 158
265 AGCAGA.....GTAGAAGCACAAAGTTTCAGAGAGATAGGGAAGTTAA 308
|||||
159 AsnArgGlnAlaValLysAlaGlnIleAlaThrGlyValGlnSerTyrI 175
309 AGAGCTCTTCGAGCTTCTCCAAAGGATTCGGTCCGACAGCTCGCT 358
175 eaSnThrMetGlyHisGluProAsnGlyIleTrpLeuProGluCysAla 192
359 ATGACCCGATAATCCCTGCCATCTACTAGGACAAACGGTTATGATATCT 408
192 yrThrTyrGlyIleAspAsnIleLeuSerGluPheGlyIleLysTyrPhe 208
409 TTGCCGACGGGAGCGGATGCTTTCTCAGCTCATCTCAACTCGCGCAT 458
209 IleSerGluGlyLysAlaIleAspTyrAlaSerPro..... 220
459 AAAGCCAATTAACCCGCTCTATCCACACCTTATAAAGGCCCAAGGAAA 508
221 .....LysProMetTyr..... 224
509 AGCGCTTAGGTACATCAGCTATCTCTTGGTCTCAGGAGCTTAGGAAG 558
224 ..... 224
559 GCGATAAAGCTCGCTTTTGAAGGTAAGGTAACGCTAAAGGCAGTCAAAGA 608
224 ..... 224
609 CATCGAAGCGGTACCCGTTTGGTGGCCGTGAACACGGCTGTAAATGCTCG 658
225 .....GlyThrAsnThr..... 228
659 GCATCGAAGGCTTCCTCTATGATCCTAAGAAAGTGCGAGCTGGATA 708
229 .....ProIleAlaAlaProSerGlyValCysAlaPheGly 240
709 GAGGACAAGACAAACATCTTCTATACGCGCACCGATATAGATTCATTGG 758
241 ArgAspMetAspSer.....SerTyrGlnValTrpSerAspPheMetG1 255
759 C.....TATAGGACATTCGAGCT 778
255 yTyrProGlyAspPheAsnTyrArgGluPheTyrArgAspIle...GlyP 271
779 ACAGAATGAGTGTGAGGATTATTAGAGGTTATAGACGAGCTCAACTCG 828
271 heGluLeuProMetGluTyrIleLysProTyrIleAsnGlu.....Asn 285
829 GAAGTGGCTTCCTCCACAGCTGAAGCAC.....AG 860
286 GlyIleArgIleAspThrGlyPheLysTyrTyrLysIleThrGlyAsnSe 302
861 TGAAGGAGGAG...CTCTACTTACGGACTTCGAGT.....TGGG 895
302 rGlyGluLysGlyIleTyrAsnArgGluAsnAlaMetLysLysValTyrP 319
896 CACCAGATAAGAGCTTGAGGATATGGAGAGGACGACGAGGGGAAAC...GCA 942

```

```

319 LUHisAlaSerHisPheAlaSerCysArgHisAspGlnIleAsnAlaAla 335
943 AGACITATATG ..... 954
336 AlaAlaAsnMetAspLysProProIleIleThrCysProTyrAspThrG1 352
954 ..... 954
352 uLeuTyrGlyHisTrpTrpPheGluGlyProAspPheIleAsnAlaPheI 369
955 .....CTGTCTACATATGAGGGGCAACTC 981
369 leargLysSerIaGluAspTrpThrSerTyrGluLeuIleThrProThr 385
982 GCCCTTTTACCGCAGACAGCATCAGGGGATGGAGCCCTCCCTCGA 1031
386 GluTyrLeuLysAsnAsnSerMetValGlnCysSerProSerPro.. 401
1032 GAGGAGCTGGATGCTTCCGGCGATATATACCATGGAGGGGTGAAA 1081
402 .....SerserTrp...GlyGluA 407

1082 ATGGGGAA 1089
407 snGlyAsp 409

seq_name: pir2:E75206

seq_documentation_block:
alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: E75206
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: E75206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-655 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CAB49100.1; PID:9545760
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: amYA; PAB0118
C:Superfamily: Dictyoglonus thermophilum amylase A

alignment_scores:
Quality: 145.50 Length: 401
Ratio: 0.808 Gaps: 22
Percent Similarity: 44.888 Percent identity: 22.444

alignment_block:
US-09-886-400-3 x E75206 ..
Align seg 1/1 to: E75206 from: 1 to: 655

73 GAGAGGCATACATCCAGCTCGAGACACTGATTAAAGAAAGAAATTC 122
|||||
28 GluLysAlaTyrArgProPheLeuGluIleLeu.....GluGluTyrPr 42
|||||
123 T.....TTTGGCTCAACATACCGGCTATACCTTAAGTTTCCTCC 163
|||||
42 oAsnMetLysValAlaIleHisIleSerGlyIleLeuValGluTrpLeuG 59
|||||
164 CGAAG.....GATATTAGACCTCGTTAAAGGGGATCCGAGT 204
|||||
59 luGluAsnLysProAspTyrIleAspLeuLysSerLeuValArgLys 75
|||||
205 GACCTGATAGATATACGACGAGCTACACGACGAATACTCCCTC 254
|||||
76 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAlaI 92

```

```

255 CCTCCGCTTAGCAGTAGAAGCACAAAGTTTCAGAGAGATAGGAAGTTA 304
|||||
92 aIleProLysGluAspArgLeuGluGlnIleTyrLeuLeuLysGluTrpA 109
|||||
305 AGAAGAGCTCTTCGAGCTTCTCCAAAGGAGGATTCGGTCCAGAGCTC 354
|||||
109 laLysLysIle...GlyTyrAspAlaLysGlyLeuTrpLeuThrGluArg 124
|||||
355 GCCTATGACCCGATATCCCTCCCATCTGAAGGACACACGTTATGAGTA 404
|||||
125 ValTrpGlnProGluLeuValLysThrLeuArgGluAlaGlyIleGluTyr 141
|||||
405 TCTATTCCGCGACGGGAGGGAGTGCTTTTCTCAGCTCATCTCAACTCGG 454
|||||
141 rValValValAsp.....AspTyrHisPheMetSerA 152
|||||
455 CG...ATRAAGCCAAATTAACCGCTCTATCCACACCTTATAAAGGCCCAA 501
|||||
152 laGlyLeuSerLysAspGlnLeuPheTrpProTyrTyrThrGluAspGly 168
|||||
502 AGGGAA.....AAGCGCTTTAGGTAC..... 522
|||||
169 GlyGluValIleThrValPheProIleAspGluLysLeuArgTyrLeuI1 185
|||||
523 .....ATCAGCTATCTCCTCGTCTCAGGG 547
|||||
185 eProPheArgProValAspLysValIleSerTyrLeuHisSerLeuAlaS 202
|||||
548 AGCTTAGGAAGCGGATAAAGCTCGTTTTTGAAGTAAGTAACGCTAAAG 597
|||||
202 erGluAspGluSerLysValAlaValPheHis..... 212
|||||
598 GCAGTCAAGACATCGAAGCCGTACCCCTTTGGGTGGCGCTGACACGGC 647
|||||
213 .....AspAspGlyGluLysPheGlyIleTrp..... 221
|||||
648 TGTAAATGCTCGGCATCGGAAGCTTCTCTTATGAATCCTAAGAAAGTGG 697
|||||
222 .....ProMet.....Thr 225
|||||
698 CGAGCTGGATAGGACACAGGAC..... 720
|||||
225 yrgIuTrpValTyrGluLysGlyTyrLeuArgGluPhePheAspArgVal 241
|||||
721 .....AACATTCTCTATACGGCACCGCATATAGAG... 750
|||||
242 SerSerAspLualAlaIleAsnIleMetLeuTyrSerGluTyrLeuGlnLy 258
|||||
751 .....TTCATTGGCTATAGGACATTCGACGGCTAC..... 780
|||||
258 sPheLysProLysGlyLeuValTyrLeuProIleAlaSerTyrPheGluM 275
|||||
781 .....AGAATGAGTGTGAGGA 798
|||||
275 etSerGluTrpSerLeuProAlaGlnGlnAlaLysLeuPheValGlu... 290
|||||
799 TTATTAGAGGTTATAGACGAGCTCAACTCGGAAGTGTGCTTCCCTCAGA 848
|||||
291 PheValGluLysLeuLysGluLeuAsn..... 299
|||||
849 GCTGAAGACACATGGAAGGACCTCTACTTACGGNACTTCGAGTTGGGCAC 898
|||||
300 ...MetPheGluArgTyrArgValPheValArgGlyGlyIleTrp.... 313
|||||
899 CAGATAAGAGCTTGAGGATATGAGAGGAGGACGAAGGAAC..... 939
|||||
314 .....LysAsnPhe...PheTyrLysTyrProGluAlaAsnTyrMetHis 327
|||||
940 GCAAGACTTTAATATGCTGTCTACAAATATGAGGGCGCAACTCGCCCTTT 989
|||||
328 LysArgMetLeuMetLeuSer.....ArgLeuLe 337

```



C;Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 21-Jul-2000  
 C;Accession: A49512  
 R;Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, C.E.  
 J. Biol. Chem. 268, 24402-24407, 1993  
 A;Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus. C  
 A;Reference number: A49512; MUID:94043280  
 A;Accession: A49512  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-649 <LAD>  
 A;Cross-references: GB:L22346; NID:g347939; PIDN:AA72035.1; PID:g347940

C;Genetics:  
 A;Gene: amyA  
 A;Start codon: GTG  
 C;Function:  
 A;Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds  
 A;Pathway: glycogen/starch degradation  
 C;Superfamily: Dictyoglomus thermophilum amylase A  
 C;Keywords: glycosidase; hydrolase; polysaccharide degradation

alignment\_scores:  
 Quality: 141.50 Length: 391  
 Ratio: 0.773 Gaps: 23  
 Percent Similarity: 46.803 Percent Identity: 22.762

alignment\_block:

US-09-886-400-3 x A49512

Align seg 1/1 to: A49512 from: 1 to: 649

```

73 GAGAGGATACATCCAGTCATCGAGACATGATTAAAGAGAAATTC 122
|||||:||||| ||| :||||| ||||| ||||| ||||| ||
29 GluLysCysTyrTrpProPheLeuGluThrLeu.....GluGluTyrPr 43
123 T.....TTTGGGTCACATAAGCGGTATACCTTAAAGTTCCTCC 163
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
43 oAsnMetLysValAlaIleHisThrSerGlyProLeuIleGluTrpLeuG 60
164 CG.....AAGATATTATAGACCTCGTTAAAGGGGCAATCCGAGT 204
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 InAspAsnArgProGluTyrIleAspLeuLeuArgSerLeuValLysArg 76
205 GACCTGATAGATTAATCGGACGAGGTACACGACGCAATACCTCCCT 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||
77 GlyGlnValGluIleValAlaGlyPheTyrGluProValLeuAlaSe 93
255 CTTCCCGCTTAGCAGAGTAGAAGCAAGTTTCAGAGATAGGGAAGTTA 304
||||| ||| :|||||:||||| ||| :|||
93 rIlePro.....LysGluAspArgIleGluGlnIleArgLeuMetL 107
305 AGGAA.....GAGCTCTCGAGCTTTCTCCAAAGGGATTCTGGCTCCA 348
||||| ||| :|||||:|||||:|||||:|||||:|||||
107 ysGluTrpAlaLysSerIleGlyPheAspAlaArgGlyValTrpLeuThr 123
349 GAGCTCGGCTATGACCCGATAATCCCTGCCATCTGAAGACACGGTTA 398
||| :|||||:||||| ||| :|||||:|||||:|||||
124 GluArgValTrpGlnProGluLeuValLysThrLeuLysGluSerGlyI 140
399 TGATGATCTATTGCGCGAGGGGAGCGATCTTTCTCAGCTCATCTCA 448
|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 eAspTyrValIleValAsp.....AspTyrHisPheM 151
449 ACTCGCGGATAAAGCAATTAACCGCTCTAT...CCACACCTTATAAAG 495
||||| ||| :|||||:||||| |||||
151 etSerAlaGlyLeuSerLysGluGluLeuTyrTrpProTyrThrGlu 167
496 GCCCAAGGGAA.....AAGCGCTTTAGGTA 521
|||
168 AspGlyGlyGluValIleAlaValPheProIleAspGluLysLeuArg.. 183
522 CATCAGCTATCTCTTGTCTCAGGAGCTTAGGAAGCGCATAAAGCTC. 570
|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

184 .....TyrLeuIleProPheArgProValAspLysValLeuGluTyrL 198
571 .....GTTTTCAGAGTAAGTAACGCTAAAGCGAGTC.....AAA 606
198 euHisSerLeuIleAspGlyAspGluSerValAlaValPheHisAsp 214
607 GACATCGAAGCGGTACCGTTTGGTGGCCGTGAACACGGCTGTAATGCT 656
||| ||| :|||||:||||| |||
215 AspGlyGluLysPheGlyIleTrpProGlyThrTyrGluTrpValTyr.. 230
657 CGGCATCGAAGGCTTCCTCTATCAATCCTTAAGAAAGTGGCGAGCTGGA 706
231 .....GluLysGlyTrpL 235
707 TA.....GAGGACAAGGACACATCTT 729
||:|||||:|||||:|||||:|||||:|||||:|||||
235 euArgGluPhePheAspArgIleSerSerAspGluLysIleAsnLeuMet 251
730 CTATAGCGCACCGATATAGAG.....TTCATTGGCTATAG 764
|||||:|||||:|||||:|||||:|||||:|||||
252 LeuTyrThrGluTyrLeuGluLysTyrLysProArgLysLeuValTyrLe 268
765 GCACATTGCAGCTACAGATGAGTGTGGGATTATTAGAGCTTATAG 814
|||||:|||||:|||||:|||||:|||||:|||||
268 uproIleAlaSerTyr.....PheGluMet.... 276
815 ACGAGCTCACTCGGAAGTGCCTTCCC..... 843
|||||:|||||:|||||:|||||:|||||:|||||
277 .....SerGluTrpSerLeuProAlaLysGlnAlaArgLeuPhe 289
844 .....TCAGAGCTGAAGCACAGTGA.....AGGGA 869
290 ValGluPheValAsnGluLeuLysValLysGlyIlePheGluLysTyrAr 306
870 GCTCTACTACGAGCTTCGAGTTGGCCACCAGATAAGAGCTTGAGGATAT 919
|||||:|||||:|||||:|||||:|||||:|||||
306 gValPheValArgGlyGlyIleTrp.....LysAsnPhe...PheT 319
920 CGAGAGAGAGAGAGGGAAC.....GCAAGACTTAATATGCTGCTCC 960
|||||:|||||:|||||:|||||:|||||:|||||
319 yLysTyrProGluSerAsnTyrMetHisLysArgMetLeuMetValSer 335
961 TACAATATGAGGGGAACTCGCCCTTTAGCCGAGAACACGCGATGCAAG 1010
|||||:|||||:|||||:|||||:|||||:|||||
336 LysLeuValArgAsnAsn..... 341
1011 GGGATGGGAGGCCCTCCTCGAGAGAGGCTGGATGCTTCGGGCGGATAT 1060
||||| ||| :|||||:|||||:|||||:|||||
342 .....ProGluAlaArgLysTyrLeuLeuArgAlaGlnC 353
1061 ATAACGAT.....TGGAGGGGT 1077
|||||:|||||:|||||:|||||:|||||:|||||
353 ysAsnAspAlaTyrTrpHisGly 360

```

seq\_name: pir2:D71011

seq\_documentation\_block:

hypothetical protein PH1386 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000

C;Accession: D71011

R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haino, Y.; Yamamoto, S.; S.

M.; Ofuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og

DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A;Reference number: A71000; MUID:98344137

A;Accession: D71011

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-560 <RAW>

A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30492.1; PID:g3257809

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenB.



C:Genetics:  
A:Gene: PH1386  
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

```
alignment_scores:
  Quality: 130.50      Length: 234
  Ratio: 1.186        Gaps: 12
  Percent Similarity: 47.009      Percent Identity: 25.641

alignment_block:
US-09-886-400-3 x D71011
..
Align seg 1/1 to: D71011 from: 1 to: 560

40 GAATCCCAAGACGCGAAATCCCAAGGTATAGAGAAGGCATACATCCC 89
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 GlutrylLeuysArgGluPheGluLysTyrMetGluArgLysLeuysSe 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 AGTCATCGAGACACTG.....ATTAAAGAGAA 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 rMetGluGluAspGluArgPheLysAspGluLysLeuArgGluAlaI 104
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 TTCCTTTGGGCTCAACATACACGGGCTATACCTTAAAGTTCCTCCGAAG 168
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 LeAsnPhe.....MetileGlyTyrPhe.....Lys 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 GATATTATAGACCTC.....GTTAAGGGGCGATCGGAGTGACCT 209
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 AspValTyrSerTyrTrpLysSerileAspGlyAsnIleLeuGlyLysPh 129
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 G.....ATAGAGATAATCGGAACGAGCTACA 235
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129 eArgGluLeuGluAspGluGlyTyrValGluValIleThrSerAlaAlaT 146
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 CGCAGCAATACCTCCCTCCTCCCGCTAGCAGA...GTAGNAGCACAA 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 hrHisGlyTyrLeuProleuGluYargAspGluAlaIleGluAlaGln 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 GTTCAGAGAGATAGGGAAGTAAAGGAGAGCTCTCGAGCTTTCTCCAAA 332
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 LeuLeuAsnGlyIleLysValTyrGluLysTyrPheGlyArgLysProAr 179
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
333 GGGATTCTGGCTCCGAGCTGCCCTATACCGGATA..... 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 gGlyIleTrpLeuProGluCysAlaTyrArgProAspGlyLeuTrpLys 196
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 .....ATCCCTGCCATACG 384
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 erProSerThrGlyValLysTrpArgLysGlyIleGluHisPheLeu 212
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 AAGCACACGGTTATAGTATCTATTCCGGCGAGGGGAGCGATGCTTTT 434
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
213 LysLysPheGlyIleGluTyrPhePheValGlu..... 223
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
435 CTCAGCTCATCTCAACTCGCGGATAAAGCAATTAACCGCTCTATCCAC 484
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 ....SerHisLeuIleAspLys...GlyProValSerLeuArgTyrGlyA 238
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
485 ACCTATAAGGCCCAAGGGAACCGGTTAGTATCATCATCATCATC 534
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 snIleLeuProAla...LysThrLysArgSerThrLeuArgProTyrPhe 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
535 CTTCGCTCTCAGGAGCTTAGGAAGGGATAAGCTGTTTTCGAAGGTAA 584
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 .....LeuLysAsnGlyIleAlaValPheAlaArgAsnAr 265
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
585 GGTAACGCTAAGGCGAGTCAAGACATCGAAGCCGTCACCGTTGGTGG 634
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 gGluThr.....GlyIleGlnValTrpSerA 274
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
635 CC 636
|||
```

274 la 274

seq\_name: pir2.G83826

seq\_documentation\_block:  
hypothetical protein BH1415 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
R:Accession: G83826  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: AB3650; MUID:20512582; PMID:11058132  
A:Accession: G83826  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-923 <STO>  
A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05134.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH1415

alignment\_scores:  
Quality: 128.00 Length: 482  
Ratio: 0.612 Gaps: 24  
Percent Similarity: 43.361 Percent Identity: 20.124

alignment\_block:  
US-09-886-400-3 x G83826 ..  
Align seg 1/1 to: G83826 from: 1 to: 923  
7 GCGCTGCTCTTTCACGCGACCTCCAGTATGCGGAATCCCAAGAGCGA 56  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
7 SerLeuValLeuHisAlaHisLeuProTyrValArgHisGlnGluGluAs 23  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
57 A.....ATCCAAAGGTGCATAGAGAAGGCATACATCC 88  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
23 ArgLeuGluGluArgTrpLeuPheGluAlaMetSerGluThrTrpIleP 40  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
89 CAGTCATCGAGACACTGATTAAAGAGAAATTCCTTTGGGCTCAACATA 138  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
40 rLeuLeuTrpAlaLeuGluLysLeuProValLysHisAlaValThrIle 56  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
139 .....AC 140  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
57 SerPheThrProValMetGluMetLeuSerAspProLeuValGlnTh 73  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
141 GGCTAT.....ACCTTAAAGTTCCTCCCGAAGGAT. 171  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
73 rArgTyrLeuAsnHisLeuGluAsnThrGluGlnLeuLysLysGluG 90  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
172 .....ATTATAGACCTCGTTAA 189  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
90 LuLysArgThrAsnAspGlnArgThrGlnAsnLeuValGlnPheTyrLys 106  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
190 GGG..... 192  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
107 GlnArgTyrGluLysLeuLysAlaThrPheLeuGlnTrpAspArgAsnLe 123  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
193 .....GGCATCGCGAGTGACCTGATAGAG.....ATAA 220  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
123 uLeuIleGlyPheArgSer...LeuMetGluAsnGluGlnCysThrLeu 139  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
221 TCGAAGACGAGCTACAGCAGCGAATACTCCCTCTCTCCGCTTAGCAGA 270  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
139 etThrSerAlaAlaThrHisAlaPhePheProTyrLeuLysThrLysGlu 155  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
271 ...GTAGAACACCAAGTTCAGAGAGATAGGAGTTAAGGAGAGACTTT 317  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
156 AlaIleArgAlaGlnValArgHisGlyIleAlaCysPheGluGlnHisPh 172  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```
318 CGAGCTTTCTCCAAAGGATTTGGCTGCCAGAGCTCGCTATGACCCGA 367
172 eGlyLysProLeuGlyPheTrpLeuProGluLucysAlaPheSerProG 189
368 TAATCCCTGCCATCTAGAGCAACGGTTATGATATCTATTCGCCGAC 417
189 lyValAspArgLeuLeuPheGluGlyLeuAspGlyThrPheValAsp 205
418 GGGAGGCGATGCTTTC.....TCACCTCATCTCAACTCGGGCAT 458
206 GluHisAlaValLeuThrAlaAspProThrProHisLysGlySer.. 221
459 AAAGCCAATTAAACCGCTCTAT...CCACAC.....CTTAAAGG 496
222 .....AlaProIleTyrSerProHisGlyLeuAlaLeuPheProA 235
497 CCCAAGGGAAGGCGCTTTAGGTACATCATCTCTCTCTGCTCAGG 546
235 rgHisThrGluLeuSerAlaLysValTrpSerThrLeuGly..... 249
547 GAGCTTAGGAGCGGATAAAGCTCTTTTCAAGGTAAGGTAACGCTAA 596
250 .....TyrProGlyAspValAspTyrAr 257
597 GGCA...GTCAAGACATCGAAGCGGTACCCGTTGG..... 630
257 gGluPheTyrArgAspIleAlaTyrAspArgGluTrpAspTyrIleLysP 274
631 .....GTGGCGGTGAACACGGCTGTAATCGTCCGCG 660
274 roHisValHisLysAspGlyIleArgIleAspThrGly.....LeuLys 288
661 ATCGGAAGGCTTCCTTATGAATCTTAAGAA.....GTGGCGAG 701
289 TyrHisArgIleThrGlyHisThrGluGluLysAspLeuTyrValArgL 305
702 CCGATAGAGGCAAG..... 717
305 utrPalGluLysArgValGlnGluHisAlaAsnHisPheIleGlyAlaI 322
718 .....GACAACATT 726
322 leHisHisGluIleAspGlnHisGlyGlyGlnAsnPheProTyrVal 338
727 CTTCTATACGGCAGCATATAGATTCATTGCTATAGGCATATGACG 776
339 MetValThrProPheAspAlaGluLeuPheGlyHisTrpTrpPheGlu 355
777 CTACAGATGAGTGTGAGGATTTATTAGAGTT.....ATAG 814
355 yProGluTrp...IleGluAlaLeuTyrGluGlnGlyAlaAspArgVal 371
815 ACAGCTCAACTCGGAATGTGCTTCCTCCTCAGAGCTGAACGACAGTGA 864
371 exPheIleThrProGluLeuTyrLeuGlnArgHisTyrGln...AspPhe 386
865 AGGAGCTCTACTTCGCGACTTCGAGTTGGCACCAGCAATAAGCCTT 914
387 GlnThrAlaHisValSerPheSerThrTrpGlyArgAspGlyTyrGly 403
915 GATATGAGAGGAGCAAGAGGAGCAAGCACTTAATATGCTGCTCCTACA 964
403 sValTrpLeuAsnAspHis...AsnAlaTrpMetTyrArgHisTyrHis 419
965 ATATGAGGGGCAACTCGCCCTTTTA..... 990
419 rgMetGluLysAspLeuAlaLysIleValAlaMetTyrProGlnProThr 435
991 .....GCCGAGAACGCGATGCAAGGGGATGCGAGCCCT 1025
436 ValLeuGluLysGlnAlaIleGlnGlnMetValArgGluTrp..... 449
1026 CCTGAGAGGAGGCTGGATGCTTCGGCGGATATATAACGATTGG 1071
```

```
450 .....MetLeuAlaValSerSerAspTrp 457
seq_name: pir2:A75207
seq_documentation_block:
  amylopullulanase_PAB0122 - Pyrococcus abyssi (strain Orsay)
  C:Species: Pyrococcus abyssi
  C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
  C:Accession: A75207
  R:anonymous, Genoscope
  submitted to the EMBL Data Library, July 1999
  A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome
  A:Reference number: A75001
  A:Accession: A75207
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-1362 <RAW>
  A:Cross-references: GB:AJ248283; GB:AL036836; NID:95457433; PIDN:CAB49104.1; PID:e15
  A:Experimental source: strain Orsay
  C:Genetics:
  A:Gene: apu; PAB0122
alignment_scores:
  Quality: 124.00 Length: 381
  Ratio: 0.756 Gaps: 16
  Percent Similarity: 43.045 Percent Identity: 21.522
alignment_block:
  US-09-886-400-3 x A75207
Align seg 1/1 to: A75207 from: 1 to: 1362
94 ATCGAGACACTGATTAAAGAAGAAATTCCTTTTGGGCTCAACATAACGGG 143
236 ValGluThrValLeuLysHisGlnMet...TrpLeuLeuAsnHisThr.. 250
144 CTATACCTTAAGTTCTCCCGAGGAGTATTATAGACCTCTTAAAGGG 193
251 .....PheGluGluHisGluLysIleAsnLeuLeuGly. 262
194 GCATCGGAGTGACCTGATAGAGATAATCGGAAGAGTACACGACGCA 243
263 .....AsnGlyAsnValGluValThrValProTyrThrHisPro 276
244 ATACTCCCTCTCTCCG.....CTTAGCAGAGTAGAAGCACA 281
277 IleGlyProIleLeuAsnAspPheGlyTrpTyrGluAspPheAspAlaG 293
282 AGTTCAGAGATAGGGAAGTTAAGGAAGAGTCTTC.....CAGC 322
293 nValLysLysAlaAsnGluLeuTyrLysGluTyrLeuGlyAlaGlyLysV 310
323 TTCTCCAAAGGATTCGGTGGCCAGAGCTCGCTATGACCCGATATC 372
310 alThrProLysGlyGlyTrpAlaAlaGluSerAlaLeuAsnAspLysThr 326
373 CCTGCCATCTCAAGCAACGGTTATGATGATCTATTTCGCGACGGG... 420
327 LeuGluIleLeuAlaGlnGlnGlyTrpLysTrpValMetThrAspGlnLe 343
421 .....GAGCGATGCTTTCTCAGCTCATCTCAACTCGCGGATAAAGC 463
343 uValLeuGluLysLeuGlyValProLysThrIleGluSerTyrTyrLysP 360
464 CA..... 465
360 rotPValAlaGlnPheGlyAspLysLysIleTyrLeuPheProArgAsn 376
465 ..... 465
377 HisAspLeuSerAspArgValGlyPheArgTyrAlaGlyMetAsnGlnTy 393
```



192 GGGCATC.....CGGAGTGACTGTATAGAGAATACT...GGAAACA 229  
|||::| ::|::| ::|::| ::|::| ::|::|  
85 .glyVallyslrtpargGlyAspArgPheGluPheVallyGsGlyProI 101  
::|::| ::|::| ::|::| ::|::| ::|::|  
230 GTACACGCAGCGCAATACTC.....CCCCTC 255  
::|::| ::|::| ::|::| ::|::| ::|::|  
101 lePheThrAsnValLeuAlaAspGluileAsnArgSerPrOProLys 117  
  
256 CTCGCCGCTTACAGAGTAGAGCACACAGTTCCAGACAGATAGGGAAGTTAA 305  
::|::| ::|::| ::|::| ::|::| ::|::|  
118 ThrGlnAlaAlaLeuLeuGluAlaMeGluGluLySgInleThrValoI 134  
::|::| ::|::| ::|::| ::|::| ::|::|  
306 GGAAGAGCTCTTCGAGCTTTTCTCCAAGAGGATTCTGGCTGCCAGAGCTCG 355  
::|::| ::|::| ::|::| ::|::| ::|::|  
134 uGlyGluThrPheSerLeuSerMetPrCPhePheValLeu.....Alat 149  
  
356 CQTATGACCOCGATATCCCTGCCATCTGAAGACCAACGGTTATGATGAT 405  
::|::| ::|::| ::|::| ::|::| ::|::|  
149 hrGlnAsnProLe..... 153  
  
406 CTATTGCGCGACGGGAGGATGCTTTTCTCAGCTCATCTCAACTCGCG 455  
|||::| ::|::| ::|::| ::|::| ::|::|  
154 ...GluGlnGluGlyThrTyProLeuProGluAlaGlnMeAspArgPh 169

```

seq_name: pirl:B69533
seq_documentation_block:
  methanol dehydrogenase regulatory protein (moxR) homolog - Archaeoglobus fulgidus
  C:Species: Archaeoglobus fulgidus
  C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
  C:Accession: B69553
  R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
    .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
    Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.F.; Weidman, J.F.; McDonald, L.
    Nature 390, 364-370, 1997
  A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
    Smith, H.O.; Woese, C.R.; Venter, J.C.
  A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
    A:Reference number: A65250; MUID:98049343
  A:Accession: B69553
  A:A:Status: preliminary; nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-334 <KLE>
  A:Cross-references: GB:AE001108; GB:AE000782; NID:g2589431; PIDN:AA891247.1; PID:g265067
  C:Superfamily: methanol dehydrogenase regulatory protein

```

705 GATAGAGGACAGGACACATTCCT ..... CTATAGCGCACCGATA 745  
 266 lMetAspGlyArgAspPheValIleProAspAspValLysArgValAlaVal 283  
 746 TAGAGTTCATTGGCTATAGGACACATTCGACGCTACAGAATGCTGTTGAG 795  
 283 aIGluAlaLeuAlaHisArgValIleLeuLysPheGluTyrAlaValAlu 299  
 796 GGATTATTA ..... GAGTTTATAGACAGACTCAACTTCGGAAGCTGTGCCT 839  
 300 GlyLeuArgAlaGluGluValAluGluGluIleLeuAsnSerValArgVa 316  
 840 TCCCTCAGAGCTGAAGCACACAGCTGGAAGGGAG 870  
 316 lPro ..... LysTyrGluAlaGlnGlu 323

seq\_documentation\_block:  
Hypothetical protein PAB1857 - Pyrococcus abyssi (strain Orsay)  
C:Species: Pyrococcus abyssi  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: C75120  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999

```

40  GAAATCCCAAGAGCGGAATCCCAAGAGGTGATAGAGAAGCATACATCC 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71  GluTyrlIleLysAlaIleuPheAsnArgIyMetGluArgLysIle.... 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90  AGTCATCGAGACACTGATTAAAGAAGAAATTCCTTTGGGCTCAACATAA 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86  .....LysLeuMetLysGluAspLeuLysLysAlaAspGlyLysL 99
140  CGGGGTATACCTTAAGTTCCTCCCGAAG ..... 168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99  euArgAsnAlaIleGluPheMetIleLysTyPheGluAspValTyArgL 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169  .....GATATTATAGACTCGTTAAAGGGGGCAT 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116  TyTrpSerLysIleAsnGlyAspIleIleGlyArgPheLysGlnLeuG 132
198  CGGGAGTAGCTGATAGAGATAATCGGAAGAGCTACACCGCAGCAATAC 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132  nAspGluGlyPheValGluIleIleThrSerAlaIleThrHisGlyTyrl 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248  TCCCTCCTCCCTCCGCTTAGCAGA...GTAGAAGCACAAAGTTCAGAGAG 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149  euProLeuGlyArgAspGluAlaIleAspAlaGlnIleLeuThrGly 165
295  AGGAAGATTAGGAGAGAGCTCTTCGAGCTTCTCCAAGGAGATCTGGCT 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166  IleArgValTyrlGluTyrlPheGlyLysLysProArgIyIleIlePle 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345  GCACAGCTTCGGCTATGACCCGATA..... 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182  uproGluCysAlaTyArgProAspGlyLeuTrpLysSerProSerThrG 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370  .....ATCCCTGCCATACCTGAAGGACCAACGCT 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199  lyGluIleLysTrpArgIyGlyIleGluHisPheLeuLysLysTyArgI 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
397  TATGATATCTATTCGCCGAC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216  LeuGluPhePhePheValGlu 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Result No.	Query			DB ID	Description
	Score	Match	Length		
1	161.5	8.6	633	1	AMVA_PYRHO
2	157.5	8.4	685	1	AMVL_DICTH
3	145.5	7.8	655	1	QSV298
4	141.5	7.5	467	1	AMVA_METJA
5	141.5	7.5	648	1	AMVA_PPRFU
6	121.5	6.5	653	1	MALO_PTHLI
7	121.5	6.5	659	1	NALQ_TRELI
8	112	6.0	311	1	MDH_BACSU
9	103.5	5.5	314	1	MDH_BACHD
10	101.5	5.4	705	1	RNR_AQUAE
11	100.5	5.4	312	1	MDH_BACTC
12	99.5	5.3	471	1	UDPG_PPRPY
13	98	5.2	474	1	NOEK_RHISN
14	97.5	5.2	813	1	LONZ_BORBU
15	96	5.1	312	1	MDH_BACIS
16	96	5.1	1067	1	LDNM_SCHPO
17	93.5	5.0	349	1	ID12_SYNY3
18	93.5	5.0	1018	1	SYI_ARGFU
19	93	5.0	1570	1	PK1_DICDI
20	92	4.9	428	1	Y047_METJA
21	91.5	4.9	467	1	UDPG_MUSAC
22	91	4.8	852	1	CLPB_CORGL
23	91	4.8	871	1	SLVA_AQUPY
24	91	4.8	8829	1	BIRG_HUMAN
25	90.5	4.8	476	1	UDPG_SOLTU
26	90	4.8	387	1	HEM2_SYNY3
27	90	4.8	630	1	Y235_METJA
28	90	4.8	658	1	REP_BUCAP
29	89	4.7	344	1	CHES_THEMEA
30	89	4.7	593	1	MTHS_YEARST
31	89	4.7	669	1	COTE_HUMAN
32	88.5	4.7	223	1	GTA1_RABIT
33	88.5	4.7	473	1	UDPG_HORVU

```

Query Match      8.6%; Score 161.5; DB 1; Length 633;
Best Local Similarity 23.4%; Pred. No. 4.5e-05;
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;

QY      8 GNQJVAEIPKSPKVIKAYPIVETLKEIP---FGLNITGYTLKELPK---DIIDL 61
      || :| : ||| : ||| ||| : ||| : ||| : ||| : ||| : |||
Db      17 GNFEW-----IKRAYEKAYRPLETL--EEYPNMKVAVHISGVLEWLENRPEYIDL 68

QY      62 VKGGTASDLIEIGTSTYTHAILPLLPSRVEAQQRDREYKEELFELS-----PK 111
      || :| : ||| : ||| ||| : ||| : ||| : ||| : ||| : |||
Db      69 LKSLIKKGQVEVLWVAGFVEYPLVAIP-----FEDRVQIKLSGWRKMNGYEAH 117

```

QY 112 GFWLPELAYDPIITAILKONGEYELFADGEAMLFSAHLSA-IKPIKPLYPHLIKAQREK 170  
| | | : : : | | | : : : | | : : : |  
Db 118 GLWTERVWPELVKTLREAGIEVILD-----DYHFSAGLSKEELFWPYTENGGEA 171  
| | | : : : | | | : : : | | : : : |  
QY 171 RFRY-----ISYLGLRELKKAIKLVF-----EGVTLKAVKVDIEAVPVAVNTAVML 219  
| | | : : : | | | : : : | | : : : |  
Db 172 IVFPDEKLYLIPFPVNETLEYLHSLADEDESKVAV-FHDDGEKFGAWPCTHELIV- 229  
| | | : : : | | | : : : | | : : : |  
QY 220 GIGRLPLMNPKKVASWI-----EDKNILLYXTDIEFIG-----YRDIAGY-R 261  
| | | : : : | | | : : : | | : : : |  
Db 230 -----BRGWLKEPFRIDSSDDKINLMYS---EYLSKFRPKGLVYLPIASYFE 274  
| | | : : : | | | : : : | | : : : |  
QY 262 MSVEGL-----LEVIDELSELCLPSELKSHGRELVLRTSSWAPDKSLRWREDEGN 313  
| | | : : : | | | : : : | | : : : |  
Db 275 MSEWSLPARAKLFFEIFKKL-KELNLFAYR-----IFVRGGIW---KNF-LYKYPEGN 324  
| | | : : : | | | : : : | | : : : |  
QY 314 ---ARLNLWVYNNRGEALLAENSARGWEPPLPERRLDAPRAYND---WRGENG 362  
| | | : : : | | | : : : | | : : : |  
Db 325 YMHRKRLMLS-----KLLRNN-----PTARIFVLRAQCNDAVWHGVFG 362  
| | | : : : | | | : : : | | : : : |

RESULT 2  
AMYL\_DICTH  
ID AMYL\_DICTH STANDARD; PRT; 685 AA.  
AC P09561;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Alpha-amylase 1 (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase).  
GN AMYA.  
OS Dictyoglomus thermophilum.  
OC Bacteria; Dictyoglomus group; Dictyoglomus.  
OX NGBL\_TaxID=14;  
RN [4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=H-6-12;  
RA MEDLINE=88225097; PubMed=2453362;  
RT Fukusumi S., Kanizono A., Horinouchi S., Beppu T.;  
RT "Cloning and nucleotide sequence of a heat-stable amylase gene from  
RT an anaerobic thermophile, Dictyoglomus thermophilum";  
RL Eur. J. Biochem. 174:15-21(1988)  
CC -!- FUNCTION: THIS AMYLASE IS A HIGHLY LIQUEFYING-TYPE: OLIGOMERS  
CC APPEARED AT THE BEGINNING OF INCUBATION, FOLLOWED BY A GRADED  
CC DECREASE IN THE AMOUNTS OF MALTOTRIOSE, MALTOSE AND GLUCOSE IN  
CC PROLONGED INCUBATION. IT IS HIGHLY HEAT-STABLE; THE OPTIMUM  
CC TEMPERATURE FOR ITS ACTIVITY WAS FOUND TO BE AROUND 90 DEGREES  
CC CELSIUS, BUT DECREASED TO ONLY 70% ACTIVITY AFTER 1 HOUR. NO  
CC DECREASE OF ACTIVITY WAS OBSERVED WITHIN THE SAME TIME AT 80  
CC DEGREES CELSIUS.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
CC linkages in oligosaccharides and polysaccharides.  
CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X07896; GAA30735.1; -;  
DR PIR; S00628; ALDYAT.  
DR InterPro; IPR004300; Glyco\_hydro.57.  
DR Pfam; PF03065; Glyco\_hydro.57; 1.  
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Multigene family.  
FT INIT\_MFT 0  
SQ SEQUENCE 685 AA; 81060 MW; C10941C8AS08C404 CRC64;

DB 28 EKAYPFFLEIL--EYPPNKNVAIHISGILVEWLEENRDPYIDLLKSLVRKGQVEIIVVAGF 85  
QY 79 THAILPLPLSRVEAQVQRDEVEKEELFELSPKGFWLPAYDPIPIPAILKONGEYELFA 138  
DB 86 YPVLAAAPKEDRLQEQIYLLKWKAKKI-GYDAKGLNTERWQPELKYKTLREAGIEVVVV 144  
QY 139 DGEAMFSAHNSA-IKPIKELYPHLTKAQRE-----KRFY-----ISYLL 179  
DB 145 D-----DYHMSAGLSKDLQFWPYTEDGGEVITVPIDKRLYLIPFRPDKVISYLH 198  
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVPVAVNTAVMLGIGRLPLMNPCKVASWTECK 239  
DB 199 SLASEDESKVAFH-----DDCEKEGIV-----PM-----IYEWYEK 231  
QY 240 D-----NILLVGTDIE-----FIGYRIAGY-----RMSVEG 266  
DB 232 GWLREFFRVSSDEAINIMLYSEYLOKFKGLVLYLPIASYFEMSEWSLPAQAKLFVE- 290  
QY 267 LLEVIDELNSELCLPSLKHSGRELYLTSSWAPDKSLRWREDEGN---ARLNMLSYNM 323  
DB 291 FVEKIKELN-----MFERYRVFVGGIW---KNF-FYKYPEANYMHKMLMLS--- 334  
QY 324 RGEALLAENSARDGWEPLPERRLDAPRAIYND---WRGENG 362  
DB 335 ----RLLRDNPSARF-----VLRQCNDAWYHGVFG 362  
RESULT 4  
AMYA\_METJA  
ID AMYA\_METJA STANDARD; PRT: 467 AA.  
AC Q59006;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative alpha-amylase (EC 3.2.1.1).  
GN M01611.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kierlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith C.R., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
CC linkages in oligosaccharides and polysaccharides.  
CC -1- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: U67601; AAB99631.1; -  
CC TIGR: M16111; -  
CC InterPro: IPR004300; Glyco\_hydro\_57.  
CC Pfam: PF03065; Glyco\_hydro\_57; 1.  
KW Hypothetical protein; Hydrolase; Glycosidase; Carbohydrate metabolism;

KW Complete proteome.  
SQ SEQUENCE 467 AA; 55558 MW; 40A6B1CDD4D967E CRC64;  
Query Match 7.5%; Score 141.5; DB 1; Length 467;  
Best Local Similarity 22.8%; Pred. No. 0.0011;  
Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;  
QY 14 EIPKSEIPKVIKAYIPVETLIK--EELPFGLN--ITG-----YTLKFLPKD-IIDLVKG 64  
DB 34 KLNKEVFNKANKYCIPTNELILEIDYDFKVNYSYTVFVQALEF--NDYVLDLFDK 91  
QY 65 GIASDLIIIGTSTHAILPLPL--SRVEAQVQRDEVEKEELFELSPKGFWLPAYDPI 123  
DB 92 LVKTVNVELIAETVHSLTSLFETEDFIEDIEMHRKMYKEITCFKAKVPRNTELIYNNR 151  
QY 124 IPAILKONGEYELFADG-EAMLFSAHLNSAIKPIKELYPHLTKAQREKRYIS--YLL 179  
DB 152 IAKIAKDLGKFAITEGIEKIL-----GWRSPNLYQSPDMKI 190  
QY 180 GLRELKAIKLVFEGKVTLKAVKDIEAVP-----YVAVNTAVMLGIGRLPLMNPCKV 232  
DB 191 LLRNLYSDDIGR-----FSARDWDQVPLTADKYAIWLASTPGEVINI---YMDYET 241  
QY 233 AS--WIEDKDNILLYGTDI-EFYGVR--DIAGY-RMSVEGLLEVIDELNSELCLPSLKH 286  
DB 242 GEHWKE-----TGIFEFLYLPITIAKHEHLEVVNVSEVVDLEPR----- 283  
QY 287 SGRELYLR---TSSWA-PDKSLRWREDEGNARLNMLSYNMRGELA--LLAENSARDGWEP 341  
DB 284 --GEIYVHEFATISWADTERDYSAWL---GN-KMQRTSFEKLKDIGKFIKENS NKL---- 333  
QY 342 LPERLDAPRAIYNDWR 358  
DB 334 ---KKLNKFDEIYKMYK 347  
RESULT 5  
AMYA\_PYRFU  
ID AMYA\_PYRFU STANDARD; PRT: 648 AA.  
AC P49067;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alpha-amylase (EC 3.2.1.1).  
GN AMYA OR PF0272.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=94043280; PubMed=8226590;  
RA Laderman K.A., Asada K., Uemori T., Mukai H., Taguchi Y., Kato I.,  
RA Anfinsen C.B.;  
RT "Alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
RT furiosus. Cloning and sequencing of the gene and expression in  
RT Escherichia coli.";  
RL J. Biol. Chem. 268:24402-24407(1993).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=94043279; PubMed=8226989;  
RA Laderman K.A., Davis B.R., Krutzsch H.C., Lewis M.S., Griko Y.V.,  
RA Privalov P.L., Anfinsen C.B.;  
RT "The purification and characterization of an extremely thermostable  
RT alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus  
RT furiosus.";  
RL J. Biol. Chem. 268:24394-24401(1993).  
CC -1- FUNCTION: DISPLAYS A BROAD RANGE OF SUBSTRATE SPECIFICITY, WITH  
CC THE CAPACITY TO HYDROLYZE CARBOHYDRATES AS SIMPLE AS MALTOSE.  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
CC linkages in oligosaccharides and polysaccharides.  
CC -1- PATHWAY: POLYSACCHARIDE DEGRADATION.  
CC -1- SUBUNIT: HOMODIMER.







Db 235 AASLAEMWEAILKDK-----KRVLTIAYL-----EGEYGED 267  
QY 175 -----ISYLGRLBELRAKLVPEGVTKLVKVDIAV 207  
Db 268 IYGVPTILGGDIEKIVELDTDEKATFAKIESV 304

RESULT 10  
RNR\_AQUAE STANDARD; PRT; 705 AA.  
AC 067834;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ribonuclease R (EC 3.1.-.-) (RNase R) (VacB protein homolog).  
GN RNR OR VACB OR AQ\_2046.  
OS Aquifex aeolicus.  
OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
OX NCBI\_TaxID=63363;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VF5;  
RX MEDLINE=98196666; PubMed=9537320;  
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,  
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V., Swanson R.V.,  
RT "The complete genome of the hyperthermophilic bacterium Aquifex  
RT aeolicus.";  
RL Nature 392:353-358(1998).  
CC -!- FUNCTION: 3'-5'EXORIBONUCLEASE THAT PARTICIPATES IN AN ESSENTIAL  
CELL FUNCTION. ACTS NONSPECIFICALLY ON POLY(A), POLY(U) AND  
CC RIBOSOMAL RNAs (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 SI MOTIF DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AE007792.1;  
DR InterPro: IPR002059; Cold\_shock  
DR InterPro: IPR001900; Ribonuclease\_II.  
DR InterPro: IPR003029; SI.  
DR Pfam: PF00773; RNB; 1.  
DR Pfam: PF00575; SI; 1.  
DR SMART: SM00357; CSP; 1.  
DR SMART: SM00316; SI; 2.  
DR PROSITE: PS01175; RIBONUCLEASE\_II; 1.  
DR PROSITE: PS0128; SI; 1.  
KW Hydrolase; Exonuclease; Nuclease; RNA-binding; Complete proteome.  
FT DOMAIN 515 696 SI MOTIF.  
SQ SEQUENCE 705 AA; 81537 MW; 2C08587229D1B05E CRC64;

Query Match 5.4%; Score 101.5; DB 1; Length 705;  
Best Local Similarity 16.6%; Pred. No. 2.6;  
Matches 77; Conservative 79; Mismatches 114; Indels 195; Gaps 22;  
QY 5 VFHGNLOYAEIP---KSE--IPKVIKAYIPVETLIKSE-----IPFGLNITGTLKF 53  
Db 100 VFHGVVRAKVTFFKGVKRVIIKVLKRAKDIKAVKVFEDQCYVVPDEN-AHRRLL 158  
QY 54 LPKDIDLVKGGIA-----SDLIILIG-----TSVT 79  
Db 159 SKKDCQCKLKEGEVYVLTQPTTKSPARGKVIENLGNPKFKFIADVIIRKYNLTSP 218  
QY 80 HAIL---PLLPLSRVEAQVQRDREVKEEL-----FELSPKGFV----- 114  
Db 219 EKVIKEVAIPKEIEEIKRRRLRQLCTIDPEKAGDFFDAVETLPEGYKLVVH 278

Db 235 ASLTEMVEAILKQRRVLTPIAYLEGEYGVILGVTIVGGNGLEQIELELTDVERA 294  
QY 148 HLNSAKPKIK 157  
Db 295 QLNKSVESVK 304

RESULT 9  
MDH\_BACHD STANDARD; PRT; 314 AA.  
AC Q9K849;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Malate dehydrogenase (EC 1.1.1.37).  
GN MDH OR CITH OR BH3158.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.  
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AP001517; BAB06877.1;  
DR InterPro: IPR001557; LLDH.  
DR InterPro: IPR001252; MDH\_active.  
DR InterPro: IPR000205; NAD\_binding.  
DR InterPro: IPR001236; Ldh.  
DR Pfam: PF00056; Ldh; 1.  
DR Pfam: PF02868; Ldh\_C; 1.  
DR PRINTS: PR00086; LLDHRCGNASE.  
DR PROSITE: PS00068; MDH; FALSE\_NEG.  
KW Oxidoreductase; Tricarboxylic acid cycle; NAD; Complete proteome.  
FT ACT\_SITE 153 153 PROTON-RELAY (BY SIMILARITY).  
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).  
FT ACT\_SITE 180 180 PROTON-RELAY (BY SIMILARITY).  
SQ SEQUENCE 314 AA; 33680 MW; 02D132F11E3B8E34 CRC64;

Query Match 5.5%; Score 103.5; DB 1; Length 314;  
Best Local Similarity 23.0%; Pred. No. 0.61;  
Matches 50; Conservative 32; Mismatches 74; Indels 61; Gaps 9;  
QY 10 LOYAEIPKSEIPK-----VIEKAYIPVETLIKSEIPGL-NITGYTLKFLPKDIID 60  
Db 130 MITYTKESGFPKPNVIGSGVLDIFAR---FRTEVQAEINLSVEDITGFVL----- 177  
QY 61 LVKGGIASDLIEITGTSYTHAT--LPLPLSRVEAQVQRDREVKEELFELSPKG--FWLP 116  
Db 178 ---GGHDDMVPLIRYSAGGIPLEKLLPQERDAIVERTKGGIEVGLIGNSSAYAP 234  
QY 117 ELAYDITPAIKLDNGEYVLFADGEAMFLSAHNSAIKPIKPLPLPHLIKAGREKRFY-- 174



```

Db      219 SNA--TCMKLAWAKEFAFDVAISDDADAPLLITDGTPTPLGDLGLICARLEAKLIA 277
QY      147 AHL--NSAIKPIKL-----YPHLIKAQEKRRFYISVLLGLREURKAIKLVFE--- 193
Db      278 TFIENSIGIEAASVEVVRTRVGSPYIAAMTEAVARGQRVMGF-EANGVMILGSNFSF 336
QY      194 GRVTLKAVKDTEAV-PYMWANVTAV-----MLGIGRLPLMNPKKVASWIEDKDNL- 243
Db      337 GGAISLPALPTDCVLPIIALHMAVEAKTPLSGIVAMHELPAVALSGRIENPFPDRSDALV 396
QY      244 -----LYGTDFIETGYNDAGTAFMSVEGLLEVDELNSELCPLSELKHSGREL 291
Db      397 AFLKASKANVSHLFSRIGRVAGTDVDGLRLTFEG-----GRIL 435
QY      292 YLRTSWAPDKSLRWIRE-DEGNARNLMLSYNM 323
Db      436 HIRPSGNAPE-LRCYVEADPDPAAEHLHAQGL 466

RESULT 14
LON2_BORBU
ID AC LON2_BORBU STANDARD; PRT; 813 AA.
IC OS1558;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ANP-dependent protease La homolog (EC 3.4.21.-).
GN BB0613.
OS Borrelia burgdorferi ( Lyme disease spirochete ).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
CX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Laghigia R., Tomb J.F., Fleischmann R.D., Richardson D.,
RA Dougherty B., Tomb J.F., Quackenbush J., Salzberg S., Hanson M.,
RA Peterson J., Kerlavage A.R., Adams M.D., Gocayne J.D., Weidman J.,
RA van Vugt R., Palmer N., White O., Ketchum K.A., Dodson R., Hickey E.K., Winn M.,
RA Utterback T., Wathey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL Nature 390:580-586(1997).
RC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S16; ALSO KNOWN AS THE LON
CC CC FAMILY OF ATP-DEPENDENT PROTEASES.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC
CC EMBL: AE001162; AAC66962.1; -.
CC TIGR: BB0613; -.
DR DR
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003959; AAA_subfam.
DR InterPro: IPR001887; ATP_GTP_A.
DR InterPro: IPR003111; LON.
DR InterPro: IPR001984; Lon_endopep.
DR Pfam: PF00004; AAA; 1.
DR Pfam: PF02190; LON; 1.
DR PRINTS: PR00830; ENOLAPTASE.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00464; LON; 1.
DR PROSITE: PS01046; LON_SER; 1.
KW Hydrolase; Serine protease; ATP-binding; Complete proteome.
NF_BIND 369 376 BY POTENTIAL.
FT ACT_SITE 719 719 BY SIMILARITY.

```

```

SQ SEQUENCE 813 AA; 92312 MW; 5129AA1498C5D0F6 CRC64;

Query Match
Best Local Similarity 5.2%; Score 97.5; DB 1; Length 813;
Matches 72; Conservative 52; Mismatches 113; Indels 95; Gaps 18;

QY 12 YAEIPKSE-----IPKVEKAYIPVIETLKEEIPFGINTGYT----- 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 507 YSIEKLEIAKIFLPSIIKESFLDKVIRIEDDVIFNL-IRNYTMESGVGLKRVITNL 565

QY 51 LKPLPKDII-----DLVKGGI--ASDLEIETISYTH-----AILPLPLSRVEAQVQR 97
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 566 IRLVRELLVEYSKQDIIKGNFVSPSSLHGNSLFTHPDPDIPGIYKIINNNYNYV-- 623

QY 98 DREVKEELPELSPKGF-----W-----LP-----ELAYDPIIPAILKONGVEY 135
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 624 DTEEDNLDLIKSSSGFVYGLANTYGGVLPVEATKFKKGDIIITGSLGAIMKESAQ-- 681

QY 136 LFADGEAMLFSAHLSAIPKIPLYPHLIKAQREKFRFYISYLLG-----LRELKAIK 189
   ::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 682 -LAYSIVKYSSKLNFDVRESPEIHLHPPEGATPKDPSAGITATASILSKKVPFLD 740

QY 190 LVPEGVKTLKAYKDIKAVPW---VAVTAVMLGIGR--LPLMNPKKVASIED-KNIL 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 741 LAMTGEVTLKGF-----VLPVGGIKKVLAAAYRNGISKVLKPNKDKKQYKLPVEVKDNI- 795

QY 244 LYGTDFIEFIYRDIAGYRMSVSGELLEVIDELN 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 796 ----DVKEF-----VSSLEEVFDYLN 811

RESULT 15
MDH_BACIS STANDARD; PRT; 312 AA.
AC Q59202;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN MDH.
OS Bacillus israeli.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=42189;
RN [1]
RX MEDLINE=96276423; PubMed=8694770;
RA Wynne S.A., Nicholls D.J., Scawen M.D., Sundaram T.K.;
RT "Tetrameric malate dehydrogenase from a thermophilic Bacillus:
RT cloning, sequence and overexpression of the gene encoding the enzyme
RT and isolation and characterization of the recombinant enzyme.";
PL Biochem J 317:235-245(1996)
CC -|- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X90527; CAA62129.1; -
HSSP; Q27743; ICET.
DR InterPro; IPR001557; L_LDH.
DR InterPro; IPR001252; MDH_actsite.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001236; ldh.
DR Pfam; PF00056; ldh; 1.
DR Pfam; PF02866; ldh_C; 1.

```

```

DR PRINTS: PR00086; LLDHDSGNASE.
DR PROSITE; PS00068; MDH; FALSE_NEG.
KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
FT ACT_SITE 153 153 PROTON-RELAY (BY SIMILARITY).
FT BINDING 156 156 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
FT ACT_SITE 180 180 PROTON-RELAY (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33648 MW; 64CE3BFF1B3C02D2 CRC64;

```

```

Query Match
Best Local Similarity 5.1%; Score 96; DB 1; Length 312;
Matches 36; Conservative 23; Mismatches 48; Indels 26; Gaps 6;

QY 10 LOVAEIPKSEIPK-----VIEKAYIPVIETLKEEIPFGINTGYTLKFLPKDIDL 61
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 130 MIYVVKESGFPKHRVIGSGVLDIAR---PRTFVAQE---LNL-----VKDITGF 175

QY 62 VKGGIASDLIEIIGTSYTHAI--LPLPLSRVEAQVQRDREVKELFELSPKG--FWLPE 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 176 VLGGHGDHVPVLRVSYAGGIFLEKLPKERLEAIVERTKGGGEIVNLLNGSAYAPA 235

QY 118 LAYDPIIPAILKD 130
   :|||:|||||
Db 236 ASLVENVEAIVKD 248

```

Search completed: June 2, 2002, 18:09:55  
Job time: 217 sec

